

## Another nuclear option

THE world will need a new source of energy when the oil and gas runs out: on that most are agreed, from the Friends of the Earth to the toughest champions of the nuclear industry. But how much energy? By when? And from what source? These are highly political questions, but they involve a technical one: what sources of energy are there? This question in turn, however, is not entirely apolitical. The gigantic momentum of the nuclear industry, set up to tame the atom bomb and, no doubt, to salve the consciences of the physicists ("We have known sin," said Oppenheimer, after Hiroshima), has determined the principal option: the fission of uranium nuclei. There has been too little cash for options other than nuclear and too little work on them. One thinks first of such fashionable renewable resources as wind, wave, sun, and geothermal energy. But, it turns out, it is not only renewable resources that are under-researched. There are specifically nuclear options that have passed almost unnoticed—as our feature this week shows (page 376).

The nuclear industry cannot continue indefinitely with existing reactors, for they consume only a tiny part of natural uranium—the fissile isotope  $^{235}\text{U}$ . The bulk of the uranium, consisting of stable  $^{238}\text{U}$ , is "wasted". This was clear at the beginning of the nuclear power programme, so it was natural to develop a system which would burn the  $^{238}\text{U}$ . This was the fast breeder, which lets the chain reaction in the  $^{235}\text{U}$  go so fast that some of the released neutrons convert the  $^{238}\text{U}$  to fissile plutonium,  $^{239}\text{Pu}$ . The plutonium they "bred" can in turn chain-react to produce energy. This system is not only economical in uranium—it seems on the face of it to be necessary in a principally uranium-based economy, for the energy-equivalent reserves of  $^{235}\text{U}$  that are extractable with a net yield of energy amount to only one-fifth of the oil reserves. Hence the dash of the nuclear industry towards the fast breeder reactor.

Yet there is an alternative to the fast breeder which will still conserve uranium—the thorium-uranium cycle. Perhaps it was the military need for plutonium; perhaps it was the simple logic of reaching towards the untapped 99% of uranium, the teasing  $^{238}\text{U}$  nucleus, that concentrated all attention on the fast breeder. But the Th-U cycle, which, at least according to its growing band of advocates, avoids the tremendous power density of the fast breeder core ( $\frac{1}{2}$  MW per litre, compared with 3 to 100 kW per litre for a thermal reactor), which reduces the problem of the long-lived wastes like americium and curium, which protects against

illegal diversion of fuel (it becomes highly radioactive), and which with proton accelerators to "top-up" the fissile content of the fuel rods might even avoid reprocessing altogether, has been looked at hardly at all outside some work in Canada and the US.

Pressure to work on the cycle comes from those resisting the fast breeder, and also from an unexpected source—accelerator physicists. The latter are to be distinguished from the high energy physicists who merely use accelerators. In a time of declining budgets for high energy physics the accelerator physicists are looking for someone else to sell their accelerators to; and the energy business seems a good bet. Yet they should not be dismissed for their self-interest; they are ingenious people and their proposals should be looked at carefully. Already they have initiated what may be the fastest route to fusion power (ion beam fusion). In the Th-U cycle they are offering ways of improving the efficiency of the cycle beyond what seems possible in the Canadian CANDU reactors until the cycle reaches true breeding—the creation of more fissile nuclei, from a non-fissile source, than are consumed in the reactor. It seems impossible to consider the Th-U cycle without considering the contribution the accelerator physicists can make.

Once breeding has been established in the Th-U cycle the reserves of thorium, not uranium, are the ultimate constraint. There seems to be plenty of thorium about, with major reserves in Canada, India, and the USSR. There is probably quite enough extractable with sufficient energy economy to exceed the energy reserves of the fast breeder fuelled on uranium.

Leaving aside the vexed arguments about whether we need a nuclear power source anyway, whether it is even economic, and whether its capital intensity is appropriate for the Third World (which is potentially the world's major energy consumer), it would seem at least that the thorium-uranium cycle, with accelerator backup, should be given very serious attention as part of the political energy equation. It may be, of course, that it is too late; that we will need our new sources of energy before Th-U can be developed into a commercial proposition. Or that the philosophically and—if you are of a certain leaning—politically attractive renewable resources, together with energy conservation and coal, will become practical and economic alternatives before you can say "fast breeder". In fact, given the opposition to the latter, Th-U plus accelerators may be the nuclear industry's last chance. □



# Conserving uranium without the fast breeder

The fast breeder, with all its dangers, is not the only way forward for the nuclear power programme. John Davies explains

NATURAL uranium, unaugmented by breeding, will provide a relatively small reserve of energy: uranium energy reserves amount to only one-fifth those of oil. The best-promoted nuclear solutions to this problem are the fast breeder reactor (FBR) and controlled fusion. The former is a means of stretching the uranium resources by two orders of magnitude; the latter promises virtually unlimited power from sea-water. Neither method is straightforward, although both were being discussed in undergraduate lectures twenty years ago. A commercial FBR is not yet operational and the design of a fusion reactor providing energy break-even is still in the unplanned future. Moreover FBRs, with their use of plutonium, are responsible for many of the strong, reasoned objections to nuclear power. However, few have realised that there are several other ways of breeding fissile materials that avoid fast reactors and their problems.

Breeding—creating fissile from non-fissile material—requires an intense source of neutrons. Neutrons can be produced by an FBR; or by a fusion reactor (yet to be built); or by using proton accelerators. The last option has not been widely discussed.

To make the arguments clear, I shall first describe some of the basic physics occurring in reactors, particularly to explain what "breeding" is and why it is essential. I shall then look at the objections to FBRs before examining the alternative means of breeding to demonstrate a system that I believe is practical, economically viable, and safe.

As medium-weight nuclei are the most stable, considerable amounts of energy are generated either when two light nuclei fuse together, or when a heavy nucleus splits into two smaller ones (undergoes fission). To exploit either possibility requires an energy barrier to be overcome. (If there were no energy barrier all light or heavy nuclei would by now have fused or fissioned leaving only medium-weight nuclei.) Another requirement (in fission) is a mechanism for maintaining a chain reaction—where the fission of one nucleus in turn induces the fission of others.

In fission, the energy barrier is, roughly speaking, the surface tension of the nuclear droplet. A heavy nucleus approaching fission becomes progress-

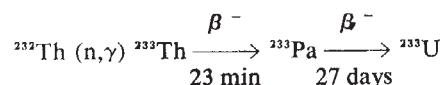
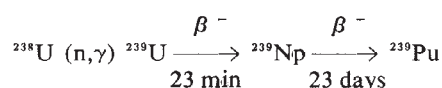
ively distorted, and the energy of the increased surface provides the fission barrier (small droplets of water are stable for the same reason). The impact and interaction of a neutron of the right energy overcomes this energy barrier. If, in turn, each resulting fission releases more neutrons (as it will, as heavy nuclei are neutron-rich) the neutrons will maintain a chain reaction with an unending succession of nuclei being induced to split.

Let us now consider fission in more detail. Thermal reactors use the fissile nuclei  $^{233}\text{U}$ ,  $^{235}\text{U}$  and  $^{239}\text{Pu}$  where the binding energy provided by the capture of a thermal neutron is sufficient to overcome the surface energy barrier against prompt fission of the resulting  $^{234}\text{U}$ ,  $^{236}\text{U}$  and  $^{240}\text{Pu}$ . Each induced fission also provides several neutrons which can then start a chain reaction. But the trouble is that the isotopes  $^{233}\text{U}$  and  $^{239}\text{Pu}$  are artificial; and  $^{238}\text{U}$  is a rare isotope of uranium, the most abundant being  $^{238}\text{U}$ .

Unfortunately, the abundant  $^{238}\text{U}$  will not do for fission, as the figure shows. This plots the effective number of neutrons per fission ( $\eta$ ) against the kinetic energy of the captured neutron; and shows the fission neutron energy spectra in thermal and fast reactors. In contrast with (say)  $^{235}\text{U}$ , the last, odd neutron in  $^{238}\text{U}$  has a smaller binding energy. Consequently only the most energetic neutron coming from fission can cause further fission by capture in  $^{238}\text{U}$  and so  $^{238}\text{U}$  cannot maintain a chain reaction.

With thermal reactors only, using natural uranium, uranium resources are insufficient to maintain a long-term nuclear energy programme. The only naturally occurring fissile isotope is  $^{235}\text{U}$ , present at 1 part in 140 of  $^{238}\text{U}$ . Moreover it is only part-burnt in thermal reactors. This leads to the idea of trying to "breed" fissile isotopes, by using the excess neutrons in the chain reaction to create fissile nuclei in otherwise non-fissile elements.

From each fission of  $^{235}\text{U}$  one neutron is required to maintain the chain reaction and 0.2 neutrons escape. Any excess over 1.2 neutrons could be used to convert  $^{238}\text{U}$  or  $^{232}\text{Th}$  to fissile  $^{239}\text{Pu}$  or  $^{233}\text{U}$  via



(Here the notation  $A(n, \gamma)B$  means nucleus A absorbs an incoming neutron and becomes nucleus B, emitting a gamma ray (a photon) in the process.

The notation  $B \xrightarrow{x \text{ min}} C$  means B decays by  $\beta^-$  emission to C with a half-life of x minutes.)

The difference between the two reactions is that the first needs fast neutrons, while the second can use slow ones. The reaction  $^{238}\text{U} \rightarrow ^{239}\text{Pu}$  occurs to a small extent in present thermal reactors (in so far as there are a few fast neutrons around) and chemical processing can enable up to 1% of all available uranium to be burnt. Even so there is probably little point in planning further thermal reactors (as there will be insufficient uranium to fuel them and existing reactors).

## Breeding in a fertile blanket

The fissile material for reactor fuel can be produced from fertile material using the excess neutrons from a reactor and the above reactions. In practice this could be done by surrounding the reactor with a 'blanket' of fertile material where the spare neutron would then 'breed' new fuel. In fact if  $\eta \geq 2.2$ , with one more neutron per fission than required to maintain the chain reaction, then breeding in a fertile blanket will maintain or increase the fissile content and allow almost all the uranium to be burnt. The figure shows that a fast reactor burning  $^{239}\text{Pu}$  surrounded by a fertile  $^{238}\text{U}$  blanket absorbing the excess fast neutrons is the best bet.

Is this really true? Much of the reasoned objections to nuclear power are consequent on the following problems of the fast breeder reactor:

- New technology—breeding around a 500 MW dustbin whose reactivity—burning rate—can be accidentally increased in micro-seconds; no artificial control is fast enough.

- Time—whether thermal reactors can produce enough  $^{239}\text{Pu}$  to fire a FBR regime whose time to breed the initial fissile content is anyway commensurate with their working life.

- The expensive, messy and dangerous chemical separation of fissile materials, in particular  $^{239}\text{Pu}$ , from spent fuel elements. This separation greatly increases the possibility of

- $^{239}\text{Pu}$  being diverted into nuclear weapons; this is further exacerbated by the considerable transporting of fissile material and by inaccuracies of assay of  $^{239}\text{Pu}$ , which can be so large as to be greater than the amounts needed for



many bombs. (This is called the "Muf"—materials unaccounted for—problem.)

● The cost—as yet not evaluated—capital, running and political—especially for trying to make them safe.

Thermal reactors on the other hand involve established technology. The figure shows that such a reactor fuelled with  $^{233}\text{U}$  has a breeding gain ( $\eta$ —2.2) greater than 1.00 i.e., as much  $^{233}\text{U}$  will be created in a  $^{232}\text{Th}$  blanket (absorbing the slow neutrons) as is burnt as fuel. Unlike  $^{239}\text{Pu}$ ,  $^{233}\text{U}$  is automatically safeguarded against illegal diversion into bomb-making by the intense  $\gamma$ -rays coming from the simultaneously produced  $^{232}\text{U}$ : additionally one can add sufficient  $^{238}\text{U}$  to make life difficult for the bomb maker but still keeping the  $^{233}\text{U}$  fissile. Such reactors avoid plutonium and the technological problems of FBRs. Also they ease the waste problem: the proportion of difficult transuranic elements (which take a long time to decay) is considerably reduced.

Thermal reactors fuelled on  $^{233}\text{U}$  can have a breeding gain of 1.00 or greater in a Th blanket. But considerable development would be necessary to make this economic. Additionally chemical re-processing of the spent fuel rods, as yet undeveloped, may be necessary.

Consider what a well established, practical and economic reactor can do with a Th-U cycle. The Canadian Deuterium Uranium (CANDU) series of heavy water moderated thermal reactors have high neutron economy and use natural uranium so there is no chemical separation. They can operate on a Th-U cycle with breeding gain as large as 0.92. Thus to use all the available thorium requires only 8% of the  $^{232}\text{Th}$  atoms to be converted to  $^{233}\text{U}$  by non-nuclear means.

Some fusion enthusiasts are lowering their expensive sights, ignoring energy output and promoting fusion reactors as intense neutron sources for fertile to fission conversion. This is not surprising! Fusion reactor design is replete with problems such that only a 2000 MW TOKAMAK has promise of creating as much energy as it consumes even with the most easily ignited  $\text{D}(\text{T}, \text{n})$   $^4\text{He}$  reaction. It is possible that the 14 MeV neutron is far more valuable for breeding than for heating, each neutron converting two or more fertile atoms into fissile ones. Even so despite the relaxed conditions for only providing neutrons fusion technology is still a dream and many regard a hybrid arrangement as combining the worst features of fission and fusion.

#### Proton accelerators

Of all the big advanced technologies, that of accelerator building can claim to have established the most consistent

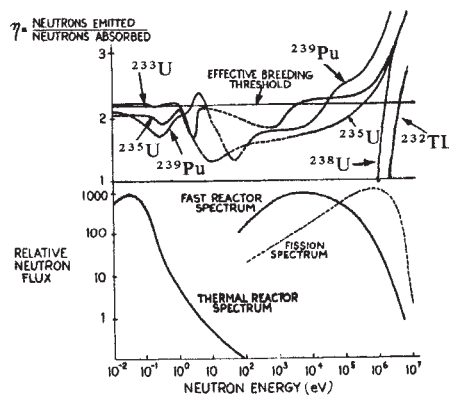
track record. The high intensity accelerator at Los Alamos is well on the way to its design beam current, the CERN SPS is working ahead of schedule within its design budget while the high energy accelerator at the Fermi Laboratory is already  $2\frac{1}{2}$  times its original design energy. At several laboratories high intensity accelerators are now replacing fission reactors as research instruments providing high intensity neutrons.

The intense neutron fluxes from such particle accelerators can effect fertile-to-fissile conversion and so produce far more energy than they consume. A 1 GeV proton interacting in a uranium target produces 60 low energy neutrons by a process known as spallation; each neutron can change one  $^{232}\text{Th}$  atom into one  $^{233}\text{U}$  atom in a fertile thorium blanket surrounding the uranium target; each  $^{233}\text{U}$  atom can fission in a CANDU reactor producing 200 MeV of available kinetic energy; of every 100  $^{233}\text{U}$  atoms so burnt, 92 will have been produced in a thorium blanket around the same or similar reactor while only 8 will have had to be produced via an accelerator. So effectively a single 1000 MeV proton produces

$$60 \times \frac{100}{8} \times 200 = 150,000 \text{ MeV}$$

This is possibly a lower estimate as there may be neutron multiplication in the blanket.

Accelerators are inherently cheaper to build than reactors and a large proportion of the accelerator operating power could be provided from heat generated in the target. The accelerator design criteria are high efficiency of radio-frequency power to beam energy conversion, minimal power usage in other parts of the accelerator (such as magnets) and high extraction efficiency.



The ratio of neutrons emitted to neutrons absorbed, in collisions between neutrons and nuclei, plotted against neutron energy. Also shown are the neutron spectra from fission directly (dotted line) and in a thermal and fast breeder reactor

Most work on meeting these requirements has involved a high intensity, medium energy, linear accelerator which is essentially an extension of the 800 MeV Los Alamos machine, now being worked up to 1 mA beam current. However, as a good example of lateral thinking, is a suggestion that these requirements can also be met by a low current, very high energy accelerator, an extension of the 400 GeV CERN SPS or the 500 GeV accelerator at the Fermilab.

A Canadian group including Fraser, Hoffman and Tunnicliffe at Chalk River are designing a 100–300 mA, 1 GeV proton linear accelerator that would breed enough fuel to supply the 8% deficiencies of a dozen, 1,000 MW CANDU reactors operating on a uranium–thorium cycle. The fertile blanket would consist of new fuel pins that would subsequently go directly to the reactor after a pause to allow for  $^{233}\text{Pa}$  decay. A sophisticated fuel management scheme is conceivable in which partially-burned fuel pins are upgraded by re-insertion in the target and then returned to the reactor.

Another suggestion, from Wilson at Fermilab, is for a superconducting 1000 GeV accelerator with an intensity of  $1\text{--}2 \times 10^{13}$  protons per second. At this energy the number of spallation neutrons is still increasing linearly with energy and so a single proton will produce 60,000 neutrons in a block of uranium. High energies and low currents mean low beam loss and efficient conversion of radio-frequency power to beam energy for protons travelling close to the velocity of light. Fermilab's superconducting energy doubler will provide 1,000 GeV protons. The intensity required for energy break-even must be increased from the present  $10^{12}$  protons per pulse by an amount that one has come to expect over an accelerator's life-time. This

neglects the effective  $\times \frac{100}{8}$  energy

multiplication obtained by breeding  $^{233}\text{U}$  for CANDU reactors.

Accelerator spallation neutron sources are being built at several laboratories as research instruments to replace fission reactors as a source of high intensity neutrons. Accelerator scientists regard upgrading spallation sources for breeding purposes as a reasonable extrapolation of existing, successful techniques. This would make the thorium–uranium cycle in thermal reactors a serious contender as a nuclear fuel source with the added bonus (to many the over-riding attraction) that objectionable material need never be removed from the fuel rods. □



# The Ganges shared out

BANGLADESH and India have agreed to share the water from the Ganges and to increase the river's flow. The interim agreement lasting five years, which the two countries signed on 5 November, states that they will share the water during the dry months from 1 January to 31 May. Estimates of the amount of water available are based on the recorded flow of the Ganges at Farakka from 1948 to 1973. The water will be released to Bangladesh for ten days at a time in three instalments each month from January to May.

The agreement also provides for water sharing during the driest period from 21 April to 30 April. At that time, water sharing between India and Bangladesh will be at the rate of 20,500 cubic feet per second and 34,500 cubic feet per second respectively. The total flow at Farakka during the period averages 55,000 cubic feet per second. The actual availability above or below the amount calculated for the period will be shared between the two countries in the same proportion. The signing of the agreement marks the culmination of protracted negotiation between the two countries which has lasted for 25 years and it is thought to be unique between riparian states.

The agreement was signed at about the same time as the Second Committee of the United Nations Environmental Programme (UNEP) constituted an intergovernmental group of experts to draft guidelines for sharing natural resources harmoniously between states. A Bangladesh representative has proposed that a conference on international rivers be organised to evolve and strengthen an institutional framework on a national, regional and international basis to deal with the problems of exploiting and developing international rivers. It is

also suggested that UNEP should be responsible for advising on the development and non-wasteful use of national resources based on sound environmental policy.

The importance of the Ganges to Bangladesh can be judged from the fact that the Ganges delta covers about 37% of the total area of Bangladesh, and contains about one-third of the total population of the country. The river provides water for drinking and other domestic and industrial uses, sustains agriculture, forestry and fisheries, and facilitates in-land navigation. The increasing demand for greater agricultural output means that water is needed throughout the year. The water available during the dry months may not be enough for the four thousand acres of paddy land dependent on the Ganges.

The task of feeding the fast growing

population of Bangladesh from the limited land and water resources is a challenging problem. Cultivating rice, the staple food of the people, is confronted with two serious handicaps—the excess of water during the monsoon and its scarcity throughout the dry months. Controlled irrigation is an alternative but expensive proposition. The Bangladesh Planning Commission, in consultation with the Water Resources Division and the Water Development Board, has proposed four pilot projects, for four different zones of the country, to develop a way of conserving water during the wet season and making it available to the farmers for the rest of the year, particularly during the lean months. The water conservancy programme will also provide a suitable environment for pisciculture and other aquatic sources of food.

**M. Kabir**

## UK Energy Commission meets

THE DESIRE of Mr Tony Benn, Britain's Secretary of State for Energy, to see a fully-integrated national energy policy came a step nearer to fruition on Monday with the first meeting in London of the newly-established Energy Commission.

The commission is an advisory body set up earlier this year to provide a regular forum for representatives of both the producers and consumers of energy to discuss long-term strategy and specific problems.

Mr Benn, speaking after Monday's meeting, said that he was aware that the Energy Commission did not represent as powerful a set of voices on

alternative as on traditional sources of energy—at least as far as research and development was concerned.

"The Government has been criticised by Sir Brian Flowers, as chairman of the Royal Commission on Environmental Pollution, and others for not paying enough attention to alternative sources of energy, and this is a criticism which we accept," he said.

One of the first requests which he had made to Sir Hermann Bondi, recently appointed chief scientist at the Department of Energy, was to look at the balance of research effort and investment between conventional and more novel forms of energy produc-

tion.

A great advantage of research in this area was that it was relatively cheap, at least in comparison with the capital-intensive investment stages. "It would be sensible for us to expand our support for research over a wide range of possibilities, and then let the development stage bid for part of the energy budget. This is an area in which I am convinced that, given the determination, we can make more efforts more quickly."

The 24-members of the commission include representatives of the energy-producing industries, relevant trade unions, various employers' organisations and consumer bodies. It has been created by Mr Benn following last year's National Energy Conference, at

Sorry, for copyright reasons some images on this page may not be available online

which the issue of energy policy was for the first time in Britain officially pushed into the arena of public debate.

According to its terms of reference, which were announced by Mr Benn in July, the commission will be concerned with two main issues: the developments of a strategy for the energy sector in the United Kingdom, and specific aspects of energy policy that arise from time to time.

On both of these, according to Mr Benn, the commission will "seek to form an agreed view", feeding into the process by which he hopes that the Government will eventually be able to publish annual reports on the energy situation. These would contain reviews of decisions taken, current prospects, and matters likely to come up for discussion.

As a first step, the Government is expected to publish a Green Paper on energy policy in Britain early in the new year. This will be based on a working document on energy policy which was prepared for the Energy Commission, and published by the Department of Energy last month.

The working document, which was welcomed by environmentalists as indicating slightly more tentative attitude

towards energy issues than previous government statements, and reduced by 10 per cent an earlier forecast of energy needs in the UK by the year 2,000, was one of the main items on the agenda of Monday's meeting.

Among the comments made, for example, was that little attention had been given to the problem of the availability of skilled manpower for the energy industries, a point which Mr Benn said would, like others made at the meeting, be taken into account in the drafting of the Green Paper.

Mr Benn expressed satisfaction, however, that in line with what he calls a "consensus approach to energy policy", the members of the commission had given general support for the working document which was to form the basis of the Green Paper. "We have hit on an energy programme for December 1977 which can claim a wide measure of public consent and understanding", he said.

He emphasised, however, that the commission was purely advisory. Although it provided a chance for the various power industries to compare notes on their future programmes, there was no suggestion that the commission should take over either the

responsibilities of the fuel industries, or those of the established Parliamentary and Governmental channels.

It remains to be seen whether this will be sufficient to maintain the confidence both of the nationalised industries, with their recently-expressed concerns about government "interference", or the trade union representatives, who would like a body such as the commission to assume a great executive responsibility.

Earlier this year Mr Frank Chapple, general secretary of the Electrical, Electronic, Telecommunication and Plumbing Union, and chairman of the TUC's Fuel and Power Industries Committee, six of whose members represent the trade union interest on the commission, suggested that there was a need for a body such as the commission to "bang heads together and work out mutually agreed policies".

Mr Chapple warned: "We are not prepared to rescue the formulation of energy policy from the short-term vagaries of the market-place only to see it taken over by bureaucrats in Whitehall and removed from the public domain."

**David Dickson**

## A second bang in the Urals

THE Kyshtym event of 1958 in many ways parallels the Tungus event of 1908. Both took place in darkest Russia. Both attracted virtually no outside attention for almost two decades. And both, once publicised, became the subject of much speculation, both informed and uninformed, with every new piece of evidence only adding to the mystery.

One of the few clear pieces of evidence to what happened at Kyshtym is given by Dr Lev Tumerman, now of the Weizmann Institute. He visited the area in 1961 and reports extensive devastation, typical of a nuclear disaster. Dr Tumerman, however, will not commit himself on the cause, stating only that it could not have been an accident in an operational power station, since at that time the Soviet Union possessed no such installation. Even in 1961, the first such station, at Beloyarsk, was still in the foundation-laying and concrete mixing stage.

Dr Zhores Medvedev, who visited the area at much the same time, is convinced that the explosion occurred in a nuclear waste store. Although the storage of waste products involves a real leakage hazard, it is difficult to see how it could cause an explosion. One Russian scientist (not a nuclear physicist), told *Nature* that he himself felt that Medvedev's explanation was quite possible, since the Soviet regime pro-

duces in its scientists an "attitude of neglect", so that proper precautions are everywhere ignored. This explanation would, however, apply equally to an explosion at a prototype reactor plant, as much as at a waste dump.

The latest step in the enigma is posed by the release of hitherto secret documents by the CIA. Fourteen documents were released; another 15 exist but are still considered too sensitive, so that, as seems typical for Kyshtym, only partial evidence is available. The involvement of the CIA itself raises a number of questions; in particular, was the notorious U-2 flight over the area

simply part of a routine Union-wide patrol, or was the unfortunate Gary Powers sent to confirm reports already received from a ground-level contact?

Much more mysterious is the fact that the CIA material mentions *two* explosions, one in 1958 and one in 1959. It seems on the face of it unlikely that two nuclear accidents could occur in successive years—even if we admit that an atmosphere of neglect makes Soviet installations peculiarly accident-prone, the fall-out from the first disaster would have surely led to the abandonment of the area, with the shut-down of any project already operating.

**Vera Rich**

## Was Tungus an astronaut?

SINCE the first expedition, just 50 years ago, to the site of the "Tungus event" of 1908, the explosion has posed a major riddle to geophysicists. The latest in a long line of theories, that of a certain Aleksei Zolotov, poses a further mystery—how, in the conditions prevailing in Soviet science, did it ever get published?

It is well-known that the publication of scientific articles in the Soviet Union is subject to strict controls. No paper may be submitted without a special certificate from the author's place of employment—a regulation which, it would appear, leaves no outlet for either the gifted amateur or the harm-

less crank. When Soviet science has, on occasion, followed its own course, unacknowledged by the outside world, this has been due to either the party bypassing normal academic procedures—as in the case of Lysenko—or else to the initiative of the topmost echelons of the scientific establishment—as with the blanket denial of ocean-floor spreading, or Snezhnevskii's theory of "creeping schizophrenia".

Recently, however, Moscow radio broadcast in both Russian and English a "new explanation" of the Tungus "enigma". The theory itself is not new to Western readers of fringe publications: the devastation, it is claimed,



was due to the explosion of an artificial "alien body", and involved nuclear reactions. Within the apparently monolithic structure of Soviet scientific publishing, its appearance, however, is startling.

There have been a number of expeditions to the Tungus area in recent years. Detailed analyses of soil and peat samples have been made, revealing, *inter alia*, a large quantity of silicate particles (diameter less than 200  $\mu\text{m}$ ) in the 1908 peat layers near the epicentre. Last March, Professor Emlen V. Subbotovich, the geophysicist, announced the findings of these expeditions: all previous theories, he said, including anti-matter, meteorite, or alien spacecraft, were wrong. The explosion was caused by a comet-like body, with density not greater than 0.9 g cm<sup>-3</sup>. His pronouncement was made with all the concomitants of a definitive pronouncement. Where, then

does Zolotov's theory fit in?

Last week, a former member of several Tungus expeditions visited London—Dr Khronid Lyubarskii, who was recently exiled for his part in the dissident human-rights campaign. Dr Lyubarskii, an astrophysicist specialising in the physics of meteors and meteorites, fully endorsed the official view that a comet was responsible for the Tungus event. According to him, Zolotov, who was described in the broadcast as a mathematician and physicist, was originally an engineer from an oil prospecting team. How he first became involved with Tungus is not clear; the impression which Dr Lyubarskii gave of him is, however, a familiar one—the self-appointed "expert" who in his own opinion knows far more than the scientist in the field. Zolotov's article, it turns out, did not appear in any scientific journal; hence no "certificate" was needed. It was

published in the monthly *Turist*—perhaps in the hope of attracting holiday-makers into the further reaches of Siberia. And for popular journals, explained Dr Lyubarskii, while there is strict political censorship, there is no control of "scientific" (or pseudo-scientific) content.

It is perhaps inevitable that genuine scientific research is surrounded by a penumbra of more or less wild speculation; indeed, this may be a necessary factor in the growth of science. One of the dangers of a monolithic state-controlled scientific establishment is that the rigid structure of the system can exclude the brilliant amateur, and also the scientist trained in one discipline, who, like Pasteur, does his most significant work in another. However wild Zolotov's theory, the fact that it can still find an official outlet in the Soviet Union is a happy omen.

Vera Rich

For years, many adult middle-class males have tried to substitute diet for exercise in warding off heart attacks. The research effort on this "diet-and-heart" question, says George Mann, "has ended in disarray." Dr Mann's conclusions were noted in *Nature* (3 November, page 2) by Dr Rivers.

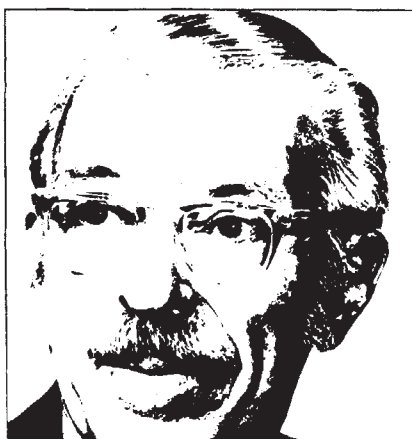
The popular dogma, which dates from around 1950, has said that diets low in saturated fats and cholesterol, but containing polyunsaturated fats (PUFAs), will produce lower blood-cholesterol levels and will lessen the risk of coronary heart disease. The idea was well-accepted; it was supported by many prominent clinicians and nutritionists. Sales of vegetable oils grew apace, especially those oils for which a slightly higher-than-average content of esters of essential PUFAs could be claimed. Big fields of thistle-like safflowers yellowed the landscape with their blossoms. Safflower seeds furnished the *cordon bleu* of vegetable oils, so rich in PUFAs that you could just feel it doing you good. The oil was hydrogenated to make margarine, even though, as George Mann points out, the hydrogenated, and hence trans, fatty acids have a hypercholesteremic effect. Butter, cream and eggs were to be shunned.

What was the effect of 25 years of these measures? By 1977, as Dr Mann points out, the results show that "a huge cohort of persons past middle age has been recruited to a futile regimen of dietary restrictions". No diet therapy has been shown effective for the prevention or treatment of coronary heart disease. Two of five dietary regimes in which vegetable

oils were featured showed an increased incidence of deaths from cancer in the treated groups.

This may be meaningless, but several experiments have shown an increase in tumour incidence in rats on diets high in polyunsaturated fats,

## Sursum corda



THOMAS H. JUKES

such as sunflower-seed oil and corn oil, as compared with saturated fats.

The vegetable oil motif was for many years the official guide to cooking and eating. Who would dare to challenge the sonorous pronouncements of medical authorities when the punishment for such heresy was incapacitation or death? We could almost feel the waxy cholesterol and tallowy fats clinging to our arterial walls after we suicidally wolfed down scrambled eggs or lamb chops. Thoughtful, health-oriented advertise-

ments gave us tasty recipes, made with vegetable oils. Sugar was added to the list of culprits; it could allegedly give rise to "hard fats" when metabolised. In any case, sugar is sweet, and therefore, sinful.

A new ascetism burgeoned and, true to type, its main objective is restoring sybaritic evil-doers to the path of dietary righteousness. Its leader is Senator George McGovern, whose committee on 'Nutrition and human needs' has issued a preposterous report on 'Dietary goals' which calls for governmental action to implement the prejudices of its writers.

The report says, among other things, that it is necessary [*sic*] to make partial substitution of polyunsaturated fat for saturated fat. Why? What about the rats with cancer? The report quotes the *New York Times* to support its thesis for reducing sugar consumption by about 40%, and speaks ominously of cola-guzzling youngsters (the little devils!) and the need for protection against "hidden sugar". Martini-guzzling adults get off without reprimand or umbrage, but we are told that monosodium glutamate "may be associated with headaches, flushes in the head and body and tingling in the spine". Since most common vegetable food proteins contain 20% or more of glutamate, there should be much tingling after a meal, and don't blame the martinis.

Avast to McGovern and his anonymous bluenoses! 'Tis the season to be jolly! I don't think they know what they're talking about, anyway, even if they have wangled a multi-million dollar appropriation to brainwash the public.



# correspondence

## In support of catastrophe theory

On 27 October we published an article by R. S. Zahler and H. J. Sussmann (page 759) critical of 'incorrect reasoning, far-fetched assumptions, erroneous consequences and exaggerated claims' in biological and social-science applications of catastrophe theory. Here is a selection of responses to the article

SIR,—Zahler and Sussmann have a basic misunderstanding of catastrophe theory: in their criticism they ignore the fundamental concept of stability. Stability lies at the root of the modern mathematical theories of dynamical systems and singularities, of which catastrophe theory is a part. The concept was introduced by Andronov and Pontryagin in 1937, and it has been greatly developed not only by Thom but also by the Russian school, notably Anosov and Arnol'd, and the American school, notably Whitney, Smale and Mather.

The importance of stability in modelling lies in the fact that if a stable model is perturbed, then its qualitative properties are preserved. Thom acknowledges the mathematical and scientific importance of stability by incorporating it into the title of his book *Structural stability and morphogenesis*, in which he first introduced catastrophe theory. On page 762 of their article Zahler and Sussmann ask the rhetorical question: "So we must ask again: what is special about a model that looks like a cusp?" and the simple answer to their question is that the cusp catastrophe is stable, whereas their figure 4 is not.

For details of their criticism Zahler and Sussmann refer the reader to a longer paper of theirs, which is not yet published. However, in the preprint of this paper, which they have circulated widely, there are major mathematical mistakes underlying their main criticisms; I explained some of these mistakes at length to Sussmann when he visited Warwick University in July this year.

Most of Zahler and Sussmann's scientific criticisms are based on misquotations, misunderstandings, misrepresentations, or quotations out of context. I give a typical example. On page 762 under the heading of 'Careless discussion of evidence' they state a number of so-called "facts", including: "Zeeman's embryology paper (*Lectures on Maths in the Life Sciences* 7, 69 (1974)),

besides being mathematically wrong, betrays the author's inexperience in embryology. For example (p. 27), Zeeman likens the embryonic neural tube to a roll of stiff paper which tries to maintain its curl. But experiment shows that cut neural tube persistently tries to unroll".

In these two sentences Zahler and Sussmann manage to misquote both Crelin and myself. Far from contradicting the metaphor, Crelin's experimental work supports it: Crelin writes (*J. exp. Zool.* 120, 561 (1952)). "The grafts of all the embryos in the rotation series showed a tendency to curl in a matter of seconds after they were severed from the brain. Therefore, if the rotation of the tectum were delayed for some reason such as the graft sticking to the forceps, the graft would curl into the shape of a ball, making it impossible to continue the operation". On page 577 figure 19, Crelin shows a photograph of the graft curled into a ball. Meanwhile on page 127 of my paper (there is no page 27) it is the underlying mesoderm, not the neural tube, that I liken to a roll of stiff paper. The burden of my discussion on that and the preceding pages concerns the forces exerted, *in vivo*, by the underlying mesoderm upon the overlying ectoderm and neural plate, before the latter has rolled up into neural tube, and at a much earlier embryonic stage than Crelin's experiment.

I now give an example of misrepresentation. When sophisticated mathematics is applied to science it is common practice to publish separately both rigorous mathematical proofs and more simplified expositions of the same material: the latter are essential if the work is to be made available to those scientists who are not expert mathematicians. A case in point is my treatment of primary and secondary waves in developmental biology. My initial simplified version, addressed primarily to biologists, is in *Lectures on Maths in the Life Sciences*, while mathematical discussions and rigorous proofs, addressed primarily to mathematicians, are in *Proc. Int. Cong. Math., Vancouver* 2, 533 (1974) and Wasserman, *G. Acta. Math.* 135, 57 (1975). The mathematical treatment uses stability with respect to a symmetry-group that is appropriate to developmental biology, namely an extension of the group of diffeomorphisms of space-time preserving the foliation by time-paths. Zahler

and Sussmann criticise the simplified version for mathematical naivety, without acknowledging the existence of the sophisticated version, although the latter has been brought to their attention long ago.

These are two examples: I could point to a hundred others. I have always found that my work has benefited from the constructive criticism of my fellow mathematicians and scientists. However, to argue in print against the determined misrepresentations of Zahler and Sussmann is both tedious for the reader and unproductive, and an adequate answer to this type of criticism is provided by my original papers *Catastrophe theory, Selected papers 1972–1977*. Nevertheless I am prompted to reply to this particular article for three reasons: firstly, their disgraceful omission of any reference to the work of Thom on biology, secondly, their implied dismissal of the fine work of Berry in physics, and thirdly, the potential harm that their article might cause to the work and the careers of other scientists who are using catastrophe theory, particularly the younger ones who have not yet established their reputations.

E. C. ZEEMAN

University of Warwick, UK

SIR,—In response to the criticisms of R. S. Zahler and H. J. Sussmann we wish to point out that contrary to misleading statements they make such as "catastrophe theory is a blind alley", several examples of genuine applications of catastrophe theory already exist in the physical sciences.

For example one of us recently studied the topological behaviour of stagnation points in two dimensional flows where, with the aid of Thom's elementary classification of degenerate critical points, a physical understanding was obtained of the complex behaviour of degenerate and non-degenerate stagnation (critical) points in a particular flow (Berry, M., & Mackley, M. R. *Phil. Trans. Roy. Soc.* 287, 1–16 (1977)). Further physical applications amenable to direct experimental test exist, some of which were discussed at the Institute of Mathematics and its Applications meeting held at University College London in May 1977.

Concerning the application of catastrophe theory to biology, we agree that some inaccurate or premature claims have been made. In particular, Zeeman's argument for the existence

of primary and secondary waves in developing embryos does not have the logical status of a proved theorem since he has not as yet published the full mathematical proof. Rather, his use of catastrophe theory in a description of differentiation gives rise to the hypothesis that such waves occur and, with the additional postulate of a temporal periodicity of state in the tissue, this application suggests how spatially periodic structures such as somites may arise.

These hypotheses have stimulated experimental investigation, which is a major purpose of model-building. Furthermore, Zeeman's treatment of differentiation has the additional virtue of providing a unitary field description of a process which is often erroneously and misleadingly described in terms of separate spatial and temporal mechanisms. In a subject such as developmental biology, which has barely begun to come to grips with its central problem of morphogenesis in terms of models, it is more important to get the correct qualitative treatment than to attempt quantitative precision.

It is far too early to decide whether or not catastrophe theory will be of major value in biology. That it provides useful and accurate descriptions of certain physical processes is now beyond question. More generally, the context for catastrophe theory is topology, and topological thinking has been of immense value in the understanding of many physical phenomena. It seems highly probable that the topological approach will prove invaluable in the study of biological processes as well, but this is an approach that can only be learned slowly, with trial and error. Zahler and Sussman have presented some valid criticisms of applied catastrophe theory, but their over-reaction is unfortunate. It leads them into exaggeration and wholesale rejection of very useful propositions.

R. BELLAIRS

University College London, UK

B. GOODWIN

M. R. MACKLEY

University of Sussex, UK

SIR,—The case in favour of catastrophe theory rests not on speculative models in the social sciences, but on successful applications to the physical sciences. In 1975 and 1976 there appeared approximately 42 papers applying catastrophe theory to physics, nine to biology, and 14 others: Sussmann and Zahler's criticisms deal almost entirely with one sociological paper, two on biology, and one model taken from two popular articles and a paragraph in a conference report. They do not hesitate to extend their conclusions to areas they have not studied: "we anticipate that

the results of an extended search (covering biology, linguistics, physics, or psychology) will be similar (that is negative)" from (Sussmann, H. J. & Zahler, R. S. *Proceedings of the 1976 biennial meeting of the Philosophy of Science Association, Chicago*, in press). Tim Poston and I have written a book (Poston, T. & Stewart, I. N. *Catastrophe theory and its applications*, Pitman, London, 477 pp.), due in print early in 1978, documenting quantitative applications in the sciences, which casts severe doubt on their conclusions. A major plank in their case—allegation of a repeated mathematical error—is refuted by Poston (*Mathematics Report*, Battelle Geneva (in press)). Their reliability may be judged by their statement: "Stewart repeats the untrue assertion that Zeeman's embryological predictions have been 'recently verified by experiment'". What I wrote was: "... with the prediction that slowing down the chemical reactions of the primary wave would lead to the formation of fewer somites, an effect recently verified by experiment". Which happens to be true.

Similar misinterpretations vitiate many of Sussmann and Zahler's criticisms, rendering them analogous to disproving Pythagoras' theorem by exhibiting a triangle that is not right-angled. With the exception of their discussion of the nerve impulse model, few of their criticisms are conclusive, and some are simply wrong. Others are problems of general mathematical modelling, which can usually be resolved by reference to current scientific practice. Sussmann and Zahler's charges go considerably beyond anything they have correctly substantiated.

IAN STEWART

University of Connecticut,  
USA

SIR,—It would be a pity if the strong attack by Zahler and Sussman on some biological and sociological models based on catastrophe theory, (27 October, page 759) were to mislead readers into thinking that such new and beautiful mathematics has no useful application in any science. The fact is that in this laboratory catastrophe theory is being employed in the development of new concepts, in the explanation and prediction of phenomena, and in the design of experiments, in two areas of physics.

The first is short wave optics (and quantum mechanics) where Thom's theory classifies the forms of focal surfaces (caustics) and makes it possible to give a precise description of the finest detail in the associated diffraction patterns (Arnol'd, V. I. 'Critical points of smooth functions and their normal forms' *Uspekhi Mat Nauk*

(translation: *Russian Mathematical Surveys*) **30**, 1–75 (1975); Berry, M. V. 'Waves and Thom's Theorem' *Adv. in Phys.* **25**, 1–26 (1976); Duistermaat, J. J. 'Oscillatory integrals, Lagrange immersions and unfolding singularities' *Comm Pure App Math* **27**, 207–281 (1974)). The classification describes caustics that are 'structurally stable', that is those whose forms survive perturbation. This makes catastrophe theory particularly suited to the optics of nature rather than artefacts such as microscopes and telescopes whose focussing is dominated by cylindrical symmetry.

We have made progress in understanding the optics of irregular water droplet 'lenses' (Berry, M. V. 'Waves and Thom's Theorem' *Adv. in Phys.* **25**, 1–26 (1976); Nye, J. F. 'Optical caustics in the near field from liquid drops' (submitted to *Proc. Roy. Soc.*), the fine structure of swimming pool caustics (Berry, M. V. & Nye, J. F. 'Fine structure in caustic junctions' *Nature* **267**, 34–6 (1976)), atom scattering by crystal surfaces (Berry, M. V. 'Cusped rainbows and incoherence effects in the rippling-mirror model for particles scattering from surfaces'. *J. Phys. A* **8**, 566–84 (1975)) and the statistics of twinkling starlight (Berry, M. V. 'Focusing and twinkling: critical exponents from catastrophes in non-Gaussian random short waves' (*J. Phys. A*, in press)). This last application (which has proved peculiarly resistant to more conventional forms of analysis) makes essential use of the enormous extension of Thom's classification being developed by Arnol'd (Arnol'd, V. I. 'Critical points of smooth functions and their normal forms' *Uspekhi Mat Nauk* (translation: *Russian Mathematical Surveys*) **30**, 1–75 (1975)) in the Soviet Union.

The other area is fluid mechanics, where the elliptic umbilic suggested the design of the 'sixroll mill' (Berry, M. V. & Mackley, M. R. 'The sixroll mill: unfolding an unstable persistently extensional flow'. *Phil. Trans. Roy. Soc. (London)* **287**, 1–16 (1977)), a device for studying the effects of dissolved long-chain molecules on the flow of Newtonian fluid. The mill produces a sequence of flows with fully describable instabilities, and addition of polymer is dramatically revealed by changes in the topology of the pattern of streamlines. This specialised application has now been generalised (Thorndike, A. S., Cooley, C. R. and Nye, J. F. 'The structure and evolution of vector fields and other flow fields' (submitted to *J. Phys. A*)) into a comprehensive theory of flow patterns, which has already given insight into the structure of the geostrophic wind and the move-



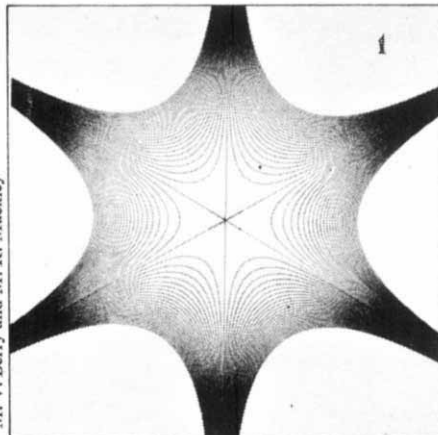
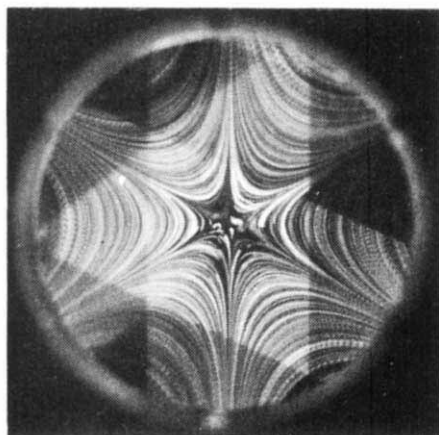
ment of ice in the Arctic ocean.

These are genuine applications of catastrophe theory; they have led to advances in our understanding of the physical systems concerned. It is important to distinguish them from illustrations of the theory, where the mathematics is employed correctly (that is to systems satisfying its axioms) but in more sophisticated derivations of results already known; elastic buckling and the mean field

theory of phase transitions fall into this category. The applications should also be distinguished from what I shall call invocations of the theory, where it is employed because of the suggestiveness of its images in the hope that its axioms might eventually be shown to apply; perhaps it is towards this area that Zahler's and Sussmann's criticisms are really directed.

MICHAEL BERRY

University of Bristol, UK



The six roll mill: experimental observations and computer simulations of the patterns illustrating elliptic umbilic critical point.

SIR,—While I do not wish to perpetuate unwarranted enthusiasm for the ability of catastrophe theory to transform the natural and social sciences, I believe that Zahler and Sussmann (27 October, page 759) have overstepped the bounds of decency in their vehement attack. Let me point out a few specific instances in which they seriously mislead the reader.

A fundamental error that they make is in the statement, "Catastrophe theorists agree that the term 'catastrophe' is reserved for certain kinds of singularity of smooth maps, seven of which have been described and classified elegantly by Thom". I can only conclude from this statement that Zahler and Sussmann have not read Thom's work. (There is no reference to Thom in the paper apart from his theorem). Thom describes the catastrophes which Zahler and Sussmann discuss as "elementary catastrophes" but repeatedly makes it clear that there are other kinds of catastrophe as well.

The authors' confusion on this point creates a straw man which they repeatedly flail in their article. For example, juxtapose the quote in the first paragraph of their paper with Thom's general definition of catastrophe and with their restricted one. The quote refers to a general approach to studying questions rather than the repeated use of a specific mathematical theorem, and makes much more sense in the context which was intended.

The section on 'Better alternatives' accuses catastrophe theory of ignoring the study of shock waves and bifurcation theory even though these are explicitly discussed in Thom's book, *Structural Stability and Morphogenesis* (Addison-Wesley, 1972), in which he lays out his theory in detail for the first time. I might add that one of the most successful applications of catastrophe theory has been to the study of shocks in a single convex conservation law (Schaeffer, D. 'Regularity theorem for conservation laws' *Advances in Mathematics* II 368–386 (1973) and Golubitsky, M. & Schaeffer, D. 'Stability of shock waves for single conservation law', *Advances in Mathematics* 16, 65–71 (1975)). In their portrayal of the scope and content of catastrophe theory, Zahler and Sussmann are simply wrong.

Let me turn to a second point. In a section entitled 'Careless discussion of evidence', Zahler and Sussmann quote Zeeman's statement: "Recent experiments by J. Cooke and T. Elsdale appear to confirm some of my predictions" (my italics). They then refute this statement with a quotation from T. Elsdale *et al.*, "we do not yet conclude that the observations here presented have confirmed Cooke and Zeeman's model to the exclusion of others" (my italics). The paper of T. Elsdale *et al.*, does indeed confirm some of the predictions of the Zeeman-Cooke model. It does not confirm the

model, but then Zeeman did not assert that it did. Zahler and Sussmann never say that these two statements contradict one another, but they clearly imply that Zeeman has made unsupported claims in his statement. This is false, and they try to mislead us into believing it.

At another point in their discussion, Zahler and Sussmann misrepresent the work of Kozak and Benham on denaturation of proteins. They assert that an essential feature of the work of Kozak and Benham is the 'delay rule' which predicts that there will be hysteresis in the denaturation-renaturation phenomenon. Yet Kozak and Benham do not rely upon this 'delay rule', and indeed use the 'Maxwell convention' throughout the second part of their three part work ('Denaturation: an example of Catastrophe II. Two-state transitions', *J. Theoretical Biology* 63, 125–149 (1976)). Zahler and Sussmann also criticise this model for predicting that the temperature denaturation curves have vertical slopes while the enthalpy change limits their steepness. Their criticism ignores the statistical discussion of Kozak and Benham which addresses the fact.

The thrust of this last criticism is also misplaced. A similar criticism could be made to the study of shocks in the solution of hyperbolic conservation laws, a phenomenon which is better understood mathematically. Real gases have viscosity, and viscosity prevents a truly discontinuous change in the velocity of the gas. Nonetheless, abrupt changes do occur. They can be modelled well by hyperbolic conservation laws which do allow discontinuous solutions. The addition of a term representing viscosity to the equations smooths out this discontinuity, and the behaviour of solutions has been studied as this viscosity term tends to zero. The conservation laws work well in giving approximations to real fluid flow. It is even the case, as we noted above, that catastrophe theory has described the shocks for the simplest (but only the simplest) conservation laws. The confusion of "the intuitive notion of 'jump' as a rapid change with the precise mathematical notion of a jump discontinuity" is not inherent to catastrophe theory, but is a common and useful approximation in many mathematical models.

I prefer to make a few remarks on the paper as a whole rather than continue to belabour specific failings. As a 'review article', Zahler and Sussmann review a single paper—their own. Their review is a summary and not a review. There are others who are more optimistic about the potential applications (and the past successes) of catastrophe theory than Zahler and



Sussmann are, but they have been ignored. Zahler and Sussmann have discovered that "catastrophe theory is a blind alley" because they have blinded themselves to anything hopeful that they might see. Surely a review of the "accomplishments of applied catastrophe theory" should not be so narrow. In particular, it should at least mention the work of J. M. T. Thompson on elastic stability (*Nature* **254**, 392-395 (1975)).

Zahler and Sussmann are also snide. They denigrate catastrophe theory for the "large number of (mostly unreferenced) publications praising each other extravagantly". This is an attempt at character assassination due to the fact that many papers are published in conference proceedings. I

think the remark is out of place—indeed I think the whole paper is out of place. It is deceptive, and it is not what is purports to be. There is much to be criticised in work which has been done using mathematical models based upon catastrophe theory, but the criticism of Zahler and Sussmann is merely mean spirited.

JOHN GUCKENHEIMER  
*University of California, Santa Cruz*  
SIR,—My paper, to which Zahler and Sussmann refer, was presented as a discussion paper at a meeting of the New York Academy of Sciences. It was directed to a general audience at a time when catastrophe theory was not widely known. Since I could not presume a detailed knowledge of the theory or of biology, I intended to give

simplified explanations of both so that I might suggest possible uses of the theory in biology.

May I also suggest that one illustration of a catastrophic jump might be the apparent discontinuity in Sussmann's opinion of Zeeman's paper (in *Towards a Theoretical Biology* **4** (1972)). In 1975 Sussmann (*Synthese* **31**, 229 (1975)) writes "in most applications what is used is not catastrophe theory as a set of results, but catastrophe theory as a conceptual framework (for an example, see Zeeman's beautiful paper in *Towards a Theoretical Biology* on heartbeat and nerve impulse)". Now, writing with Zahler he expresses a very different opinion.

A. E. R. WOODCOCK  
*Williams College, Massachusetts, USA*

## Censuring repressive regimes

SIR,—Every four years the International Union Against Cancer organises a huge international cancer congress, typically attended by perhaps 6,000 scientists. The next one is scheduled for 5-12 October 1978 in Buenos Aires, Argentina. However, in *Science* (21 October) and *Nature* (3 November, page 8) groups of scientists, including two recent Nobel prize-winners for medicine, have called on other scientists to boycott this conference because, it is claimed, "scientists, physicians, professors, journalists, intellectuals, and other (Argentinian) citizens have been arrested, imprisoned without benefit of *habeas corpus*, often tortured and sometimes executed without trial". These allegations are supported by editorials in *Nature* (**263**, 452; 1976 and **266**, 395; 1977).

We were telephoned by the press for comment on these letters, but could not comment because, like many other full time scientists, we were necessarily ignorant of conditions in most foreign countries. We therefore made contact with Amnesty International (AI), in the hope of securing some reasonably unbiased data. The 'Report of the AI Mission to Argentina' (1977) and the detailed letter which the Director of AI sent us appear to be objective and impartial, and the conditions they describe are really appalling.

Briefly, there was a military coup in March 1976. During the next six months about 2,000 Argentinians were killed by the government or (almost equivalently, it seems) disappeared. Killings and disappearances of non-violent Argentinians (a priest, for example, who said a requiem mass for a parishioner who had been killed by government agents) continue at a rate

of hundreds a month, and frequently bodies are discovered, some of people recently arrested, many with evidence of severe torture. Recent Argentinian laws explicitly forbid newspapers to report, comment on or make reference to abductions or the discovery of bodies, and also forbid propagation through any medium whatever of news or views with the purpose of lessening the prestige of the armed forces!

In the apparently impartial opinion of the Director of AI (personal communication, 25 October), "the state of affairs in Argentina is now one of the most serious in the world: Argentina has probably, since March 1976, the largest number of prisoners of conscience, disappearances and political killings in the whole of Latin America". (In most cases, according to AI, those affected were not terrorists.) However, even if, among many bad countries, Argentina is worst in this respect, it does not follow that a boycott is wise, for it is difficult to estimate the true probabilities of all its possible effects (especially if one is distracted by suppositions about the motives of the scientists who propose it or of those who oppose it).

It must be recognised by the proponents of any boycott that it is very unlikely to have much impact on the repression of non-violent citizens, while the opponents of a boycott must equally concede that there does exist a small probability that it would materially accelerate humanitarian progress. Balancing these considerations are the medical advantages of holding a cancer congress, but here a rather curious analogy exists; one single congress is very unlikely to have much impact on human cancer, but there does exist a small probability that it will materially accelerate medical progress. Moreover, because the Argentinian junta are apparently (*Financial Times*, 14 September) having difficulty,

because of their human rights record, in securing foreign investment and arms, it is uncertain which probability is smaller in this particular balance.

One must also ask, however, if scientific meetings are boycotted, where it will all end. In a related academic field, for example, the International Epidemiological Association will meet in Iran in 1979 and, overall, a large number of international scientific meetings take place in countries with governments which kill, or torture, or imprison non-violent citizens. If these several meetings were all avoided, so many different governments would thus be censured that none would thereby be singled out for special attention, and the net effect might merely be to forfeit the rather useful apolitical image which science has earned, while achieving little or no humanitarian progress. It might be practicable just to boycott those governments which persecute many scientists, but this seems rather artificial, and might merely aggravate anti-science sentiments. Moreover, the scientists who live in some of the countries concerned might much rather have a week or two of contact with the international community than suffer continued intellectual isolation. Perhaps, anomalously, boycotts do more harm than good in states with closed frontiers (unless the state would also restrict free attendance), but perhaps they are occasionally useful in states with open frontiers.

Advice and comments from other scientists, especially those who, unlike ourselves have actually worked under repressive regimes, would be timely, both about the general advisability or inadvisability of such boycotts (ignoring everybody's motives, please), and also about the specific 12th Cancer Congress in Argentina next year.

RICHARD PETO  
RICHARD DOLL  
*University of Oxford, UK*

# news and views

## Object Kowal, the most distant asteroid

from David W. Hughes

ASTEROIDS often get into the news. Only 2 years ago asteroid 1976 AA was found to be the first with a mean distance from the Sun less than that of the Earth. On 1 November this year slow-moving object Kowal was discovered and has subsequently been found to have the largest known orbit of any of the asteroids. It is orbiting the Sun between Saturn and Uranus.

Charles T. Kowal of the Hale Observatories, California Institute of Technology is carrying out a systematic survey of the outer solar system by photographing regions of the ecliptic using the 122-cm Schmidt telescope on Mount Palomar. He first spotted this strange asteroid when examining photographic plates that he had taken 2 weeks previously on 18 and 19 October. It was also discovered on the plate that had been taken by Tom Gehrels (University of Arizona) on 11 October.

Bodies in the Solar System appear to move across the sky against the backdrop of the vastly more distant, 'fixed' stars. A long-time exposure will thus produce a trail-of-light image for the Solar System objects as opposed to a point image for the stars. Also the positions of the objects change from one plate to the next. These trails can be detected by the use of microscopes.

Slow-moving object Kowal has an angular motion of only 0.0149 degrees per day, a speed scarcely greater than that of Uranus. As objects in the Solar System with near circular orbits have a mean daily angular motion proportional to  $a^{-3/2}$  where  $a$  is the semi-major axis of their orbit, this automatically means that object Kowal is in the outer reaches of the Solar System, near the orbit of Uranus. On 18 October it had a right ascension of  $2^{\text{h}} 05^{\text{m}}$  and a declination of  $+12^{\circ} 09'$ , placing it about half-way between Alpha Aries and Alpha Pisces. If the positions of the object on a set of dates are known,

its orbit can be calculated. Object Kowal has, however, moved only about three-quarters of a degree around its orbit since its discovery, thus making the orbital determination very inaccurate. Brian G. Marsden (Smithsonian Astrophysical Observatory, Cambridge) has published a preliminary calculation in Circular 3129 of the International Astronomical Union's Central Bureau for Astronomical Telegrams. This orbit is shown in Fig. 1. It has a low eccentricity, 0.031, a low inclination to the ecliptic,  $5.2^{\circ}$ , a peri-

helion distance of 15.836 AU, a semi-major axis of 16.340 AU, and a period of 66.1 years.

The object is extremely faint, having a visual magnitude between 18 and 19, this faintness being one of the reasons it has previously gone undiscovered. If it has the orbit shown in Fig. 1 its magnitude will change by only 0.5 over one period. There are thus very few telescopes capable of searching for such objects, those that can find them having to have an objective lens more than 100 cm in diameter.

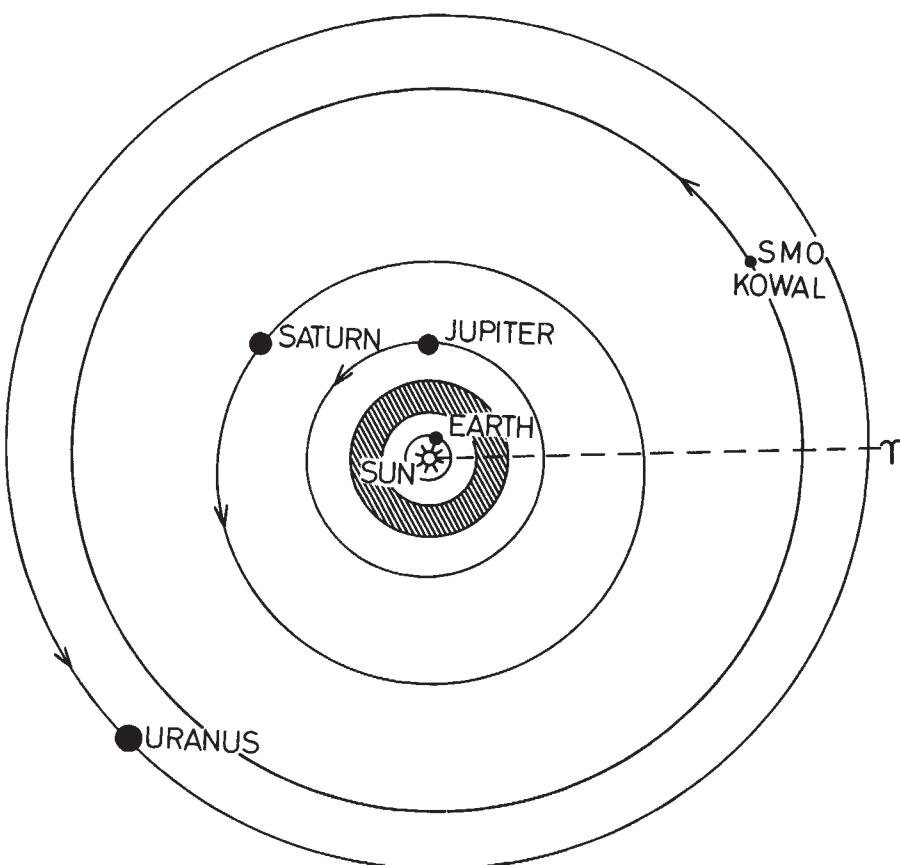


Fig. 1 The orbit of slow-moving object Kowal and the position of this new asteroid with respect to Earth, Jupiter, Saturn and Uranus. The hatched area is the main asteroid belt.  $\gamma$  is the first point of Aries.

David W. Hughes is a Lecturer in the Department of Physics at the University of Sheffield.

Even though object Kowal is faint, it is much brighter than any known comet would be at that distance. It has thus been concluded that it is an asteroid. It might even turn out to be the first of a new belt of asteroids orbiting between Saturn and Uranus. Most asteroids orbit the Sun in a broad belt between Mars and Jupiter (shown as a hatched area in Fig. 1). Some have larger orbits, Hidalgo for example has a highly eccentric orbit which approaches the orbit of Saturn. It must be remembered of course that the further away from the Sun and Earth an asteroid is, the more difficult it is to detect. The region between Saturn and Uranus could easily contain an Earth's mass worth of asteroids which have remained undetected until now.

The brightness of the asteroid can be

used to estimate the product of its albedo and its reflecting surface area. Asteroids seem to have albedoes ranging from about 0.035 to 0.27. As the albedo of object Kowal is unknown the size cannot be found accurately. It has been estimated to have a radius in the region 80–320 km. The largest asteroid, Ceres, has a radius of 380 km, and so object Kowal is definitely at the top end of the asteroidal mass scale.

Where do we go from here? More observations of object Kowal, stretching forward in time, or even lying unknown on photographic plates, will enable a much more accurate orbit to be calculated. Also now astronomers know that at least one asteroid, albeit a large one, is orbiting between Saturn and Uranus it is most likely that more will be discovered in the near future. □

## Workshop on hydrophobic effects

from Felix Franks

A workshop on Hydrophobic Effects was held at the 5th International Conference on Chemical Thermodynamics, Ronneby, Sweden, 23–26 August, 1977.

THE involvement of so-called hydrophobic effects in the maintenance of biological structures has been recognised for many years; G. S. Hartley first used the term in his analysis of micellar aggregation. The detailed nature of the environment of alkyl chains in water, and the manner in which their interactions give rise to supermolecular structures and promote the unique folding of biopolymers, are not yet too clear. Nevertheless, over the past 10 years a considerable effort has been devoted to studies of dilute aqueous solutions of alkyl derivatives. The workshop was arranged in an attempt to reach some agreement on terminology and symptoms of hydrophobic effects and to analyse those features which play a part in biochemical processes.

The question of suitable and correct standard states was discussed at length; for means of comparison, the transfer of an alkyl group from the ideal gas state to an infinitely dilute aqueous solution is the most useful process to consider, and the term 'hydrophobic solvation' was used to describe such a process. Although solution concen-

tration is commonly expressed in terms of mol fraction or molality, it was pointed out that the correct way of comparing solute environments in different solvents is through the molarity concentrations. This is also required for statistical thermodynamic treatments of solute–solute interactions, since the equations describing solution behaviour are referred to constant volume, rather than to constant pressure.

The standard thermodynamic functions of solvation which characterise the transfer of an apolar solute from the ideal gas to an infinitely dilute aqueous solution include a positive free energy which arises from a large, negative entropy, a negative volume and a very large, positive heat capacity. In fact, the limiting partial molar heat capacity of an alkane in aqueous solution is three times as large as the heat capacity of the alkane in its pure liquid state. Taken together with recent nuclear magnetic relaxation and computer simulation data, the heat capacities are consistent with the picture of a rotation of water molecules away from the alkane, with possibly slight deformations in the tetrahedral hydrogen bonding geometry, to give rise to cavities which can accommodate the hydrophobic groups. The hydrogen bonding regime which normally exists in liquid water is but slightly modified; whatever modification does take place is responsible for the negative entropy and the positive heat capacity of hydrophobic solvation.

Interactions between solvated alkyl residues can be studied at different

levels, ranging from pairwise solute effects to the complete removal of water and its substitution by a hydrocarbon. For the latter process the standard free energy of transfer is about  $-3.45$  kJ per mole  $\text{CH}_2$  groups. Intermediate between these two environments are the micelle, the phospholipid bilayer membrane and the interior of a globular protein. None of these three biologically important structures can be equated to the bulk hydrocarbon, and therefore free energy data relating to the transfer of alkyl groups from water to hydrocarbon should not be used for thermodynamic calculations of micellisation or protein unfolding. The hydrophobic pair interaction is considered to be an adequate model for enzyme–substrate interactions and for the contributions of exposed apolar amino acid residues to protein stability (or instability).

The thermodynamic functions describing such pairwise interactions are obtained as second virial coefficients from the limiting slope of the concentration dependence of the various excess functions. Recent excess volume ( $\Delta V^E$ ), heat capacity ( $\Delta C_p^E$ ) and compressibility ( $\Delta \kappa^E$ ) measurements on very dilute solutions indicate that these second virial coefficients have the opposite signs to those expected from the classical treatment of the hydrophobic interactions. Indeed,  $\Delta C_p^E$  is a very complex function of concentration and temperature. It seems that the hydrophobic pair interaction is qualitatively different from the interaction involving large numbers of alkyl chains, for example, micelle formation.

A comparison of gas and solution second virial coefficients of  $\Delta G^E$  shows that the interaction between pairs of alkyl groups in dilute aqueous solution is weaker than the corresponding interaction in the gas phase. Also, recent experiments with non-aqueous solvents (for example, dilute solutions of octane in *N*-methylacetamide or hydrazine) indicate the existence of a general solvophobic effect which differs from the hydrophobic effect in two aspects: it does not originate from a negative  $\Delta S^E$ , and it is not associated with a large positive  $\Delta C_p^E$ , both of which effects characterise the hydrophobic interaction. With rise in temperature water becomes more hydrophobic, and near  $150^\circ\text{C}$  its behaviour approaches that of 'normal' polar solvents.

So far only limited progress has been made in the molecular description of hydrophobic effects by means of suitable potential and radial distribution functions. Any useful interpretation of terms such as 'water structure making', commonly found in the literature, must await the development of credible radial distribution functions. In the meantime it is being recognised that



the 'interaction' between alkyl groups in water results primarily from repulsions between water and solute, and not from van der Waals attraction between solute molecules.

Most molecules for which hydrophobic effects are commonly invoked also possess one or more functional groups which contribute to the observed interactions in solution. In all cases such contributions of  $-\text{CONH}$ ,  $\text{CHOH}$ , and so on to  $\Delta G^\text{E}$  are negative, but only alkyl groups exhibit negative contributions to  $\Delta S^\text{E}$ . Even more important, cross interactions between alkyl and polar groups are not equal to the mean of the interactions between like groups. This finding has a bearing on calculations of protein stability in terms of amino acid side chain contributions.

Any future progress in our understanding of hydrophobic effects is not likely to come from thermodynamics, although there is still much scope for high precision measurements of excess functions in the low concentration range. Techniques which are showing promise include electron spin exchange between hydrophobic spin probes and the enhancement of nuclear magnetic relaxation by such paramagnetic spin probes. Monte Carlo and molecular dynamics simulations have also shown some promise, but in the absence of reliable potential functions it is questionable how much confidence can be attached to the results of such calculations. The same question mark hangs over recent computer simulations of protein dynamics, in which solvation and hydrophobic contributions to the stability of the folded protein were not, and indeed, could not be included. However, in view of the progress that has been achieved over the past 10 years in our general understanding of molecular interactions in aqueous solutions, the prospects for such ambitious ventures look quite encouraging. □

## Mechanisms of parasite immunity

from F. E. G. Cox

A meeting on Immunity in Parasitic Diseases: Antigens and Mechanisms of Immune Response was held under the auspices of the Institut National de la Sante et de la Recherche Medicale and the Institut National de la Recherche Agromique with Professor Andre Capron as Chairman, at Grignon, France on 5-9 September, 1977.

PARASITOLOGISTS and immunologists need to meet from time to time to share their rapidly accumulating knowledge and there can be few better places to do so than a Louis XIII chateau. Twenty review papers and four round tables, each with a dozen or more contributors, demonstrated the healthy state that the science of parasite immunology has reached and for the first time various pieces of information painstakingly obtained in different laboratories began to fall into place. The two themes that came over time and time again were first that in each infection there are probably several effector mechanisms involved in immunity and that under particular circumstances any or all of these might be involved and, second, that eosinophils, macrophages, antigen-antibody complexes and complement may be more important to parasitic infections than hitherto suspected. Against this background it was possible to look forward as well as backward.

The initiation of any immune response begins with the antigens of the parasite. Most parasite antigens studied so far have been relatively crude extracts and Ruth Arnon (Weizmann Institute of Science, Rehovot) discussed the possibilities of obtaining pure antigens which could be characterised and synthesised. For some microorganisms this possibility is a reality and for many parasites there are indications that the isolation and characterisation of functional antigens has progressed sufficiently far for the synthetic stage to be contemplated. The isolation of such antigens depends on the exploitation of their characteristics such as enzymatic activity and ability to bind to drugs and ligands and P. Pery (Grignon) showed that this was not difficult to achieve. In the case of *Schistosoma mansoni* over 60 antigens have been identified. D. Bout and his colleagues at Lille, using drugs as

ligands, have found that malate dehydrogenase is a particularly important antigen in human *S. mansoni* infections and J. P. Rotmans (University of Leiden) has shown that only one isoenzyme of malate dehydrogenase is antigenic in mice. This antigen can be purified and used in microgram quantities in immunodiagnosis of schistosomiasis. An egg antigen described by J. Hamburger (Jerusalem) is the antigen responsible for granulomas in mice but also has considerable potential in immunodiagnosis as has a circulating polysaccharide antigen isolated by Y. Carlier and his coworkers at Lille. Other purified antigens also described include one from *Nippostrongylus brasiliensis* (A. Petit *et al.*, Grignon) and one from *Trypanosoma cruzi* which, as it is shared by both the blood and culture forms of this parasite, may be very useful as a potential vaccine (L. Hudson and D. Snary, Wellcome Laboratories, Beckenham).

The effector mechanisms attracted considerable attention particularly in the case of schistosomiasis. A. Capron (Lille) drew attention to the two main ways in which schistosomula are attacked, IgG and eosinophils and IgE and macrophages. The IgG involved is IgG2a and the activity of the eosinophils is enhanced by the presence of mast cells (Monique Capron *et al.*, Lille). The importance of eosinophils in immunity to parasites was discussed in general terms by A. B. Kay (Edinburgh) and in much more detail as the cells that actually attach to schistosomula and penetrate the body wall by C. Mackenzie *et al.* (National Institute for Medical Research, London) who pointed out that these cells adhere to other helminths as well. B. M. Ogilvie (NIMR) also drew attention to the similarities in the immune responses to parasites as different as *Elmeria* and *Nippostrongylus* living in the gut.

Macrophages are obviously important in immunity to parasites and their role in the production of nonspecific mediators of immunity was discussed by A. C. Allison (Clinical Research Centre, Harrow). Nonspecific immunity to blood parasites following the administration of *Corynebacterium parvum* or BCG was also discussed by F. E. G. Cox (King's College, London), who postulated a common mechanism for homologous, heterologous and non-specific immunity to malaria in rodents; and by B. Leblanc and J. C. Salmon (Villejuif). Another nonspecific mechanism, the activation of complement through the alternative pathway, was discussed by G. Hultdt (Stockholm) in the context of *Entamoeba histolytica*.

Immunodiagnosis and vaccination are the two main objectives behind immunological studies and the prospects



### A hundred years ago

THE chemists of Berlin have been occupied lately in analysing the wares of the wine merchants, and no little excitement has been caused by the discovery that the entire stock of one of the largest houses dealing in wines for medicinal purposes, consisted entirely of artificially prepared mixtures of spirit and sugar solutions, flavoured with various herbs.

From *Nature* 17, 29 November, 1877.

F. E. G. Cox is Professor of Zoology at King's College, London.

for vaccination were discussed by D. S. Rowe (World Health Organisation, Geneva) but it is too early yet to assess the potential of purified or crude antigens as vaccines against the more important parasitic diseases. The successful vaccination of cattle and sheep against *Schistosoma bovis* using irradiated whole larvae was described by M. G. Taylor (St Albans) but a more cautionary story was told by G. M. Urquhart (University of Glasgow) who discussed the success of the irradiated *Dictyocaulus viviparus* vaccine and the lack of success with a similar *Ancylostoma caninum* vaccine which was due not to any defect in the vaccine but to the fact that American veterinarians preferred drugs which are very effective against this worm. For most of the parasitic diseases of man effective drugs are not currently available and until they are immunological research must not be allowed to decline; those present at the Grignon meeting will ensure that this does not happen. □

## Fly's eye view of cosmic rays

from K. J. Orford

AN interesting and elegant new experimental technique, designed to take 'pictures' of the tracks in the atmosphere of very high energy cosmic rays, has borne its first fruit. The results of a pilot experiment (Bergeson *et al.* *Phys. Rev. Lett.* **39**, 847; 1977) have demonstrated that such tracks may be observed using the very weak atmospheric nitrogen fluorescence light. Its significance is that the effective size of detectors of the very rarest cosmic rays has been increased by this technique to the volume of the atmosphere which can be seen from a point on the ground.

The main aim is to measure the energy spectrum of cosmic rays up to energies difficult to attain with any other technique. A continuous spectrum of energies is found to come from all parts of the sky up to an energy of about  $10^{20}$  eV (about 16 J) per particle. The cosmological problems centre on the sources of these particles, whether they are local or universal, and on how much such prodigious energies can be given to individual particles. If, for example, these cosmic rays are universal, then a limit is placed on the energy spectrum. At an energy near  $6.10^{19}$  eV a proton will interact strongly

with the 2.7 K universal microwave background, and will lose a large amount of energy. The spectrum will then rapidly fall in the region of  $10^{20}$  eV. There are indications from current experiments that it does not (Cunningham *et al.*, *15th Int. Conf. on Cosmic Rays*, Plovdiv 1977), at least up to nearly  $10^{20}$  eV.

Just to detect the arrival of these cosmic rays is very difficult, mainly as a result of the small flux. A cosmic ray of energy greater than  $10^{20}$  eV falls on a square kilometre somewhere between once per year and once per century, such is the present uncertainty. Such a low rate is detectable only because the cosmic ray particle causes a shower of up to  $10^{11}$  electrons and other particles to build up along its track as it passes through the air. This shower has a lateral extent of a square kilometre or more, enabling particle detectors to be widely spaced on the ground to provide samples of the shower particles. For about 20 years, ever larger arrays of such detectors have been deployed around the world to sample large showers.

Some of the best information to date has come from the UK array at Haverah Park, near Harrogate. This has a sampling area of 12 square kilometres and has recorded showers of energy up to  $10^{20}$  eV. However, there is a limit to the size which such arrays can be made, the cost increasing with the area covered.

The new technique uses the whole atmosphere visible from any point as a detector. It exploits the fact that most of the energy of the primary cosmic ray does not reach the ground, but is dissipated as ionisation in the air. About 0.05% of this energy is released as nitrogen fluorescence light which produced a very short ( $\mu$ s) flash of light along the cosmic ray track. A detector designed to measure this flash was first proposed in 1965 (Greisen *Int. Cosmic Ray Conf.*, London, 1965). However, the technical problems are severe. The light output is very small and difficult to detect against the background light of the night sky (each electron in the shower gives only four fluorescence photons per metre of track).

The pilot experiment of Bergeson *et al.* involved the use of concave mirrors of 1.5 m diameter, each focusing a small section of sky onto the faces of 12 photomultipliers. Each tube views an area of sky about  $6^\circ$  across. Showers detected by a conventional particle array at Volcano Ranch, New Mexico, were compared with the images seen by the mirror assemblies and found to be in agreement. Following this, a full-scale experiment is being constructed on a mountain top in Utah, consisting of 67 mirror units. Thus a whole sky

image is obtained with a resolution of better than 0.007 steradians. This many-faceted experiment, early dubbed the 'Fly's Eye', should detect the highest energy showers at a distance of 50 kilometres. The effective collecting area of about 8,000 square kilometres should allow an energy spectrum to be measured with high statistical precision.

This would be enough by itself to justify the experiment, but other results of no less significance may be forthcoming. Because the track of ionisation is seen for most of its length, the point at which the primary cosmic ray particle first interacted with a nucleus in the air may be found. The observation of a large number of such points should reveal the mean free path of the primaries for interaction in air, and hence their atomic weight. The composition of high energy cosmic rays is at present an open question, and compositions other than a simple 100% proton flux bring their own problems concerned with production, acceleration and propagation from the sources to the Earth.

Finally, the mean free path of the primaries could enable the total cross section for nucleus-nucleus, and from that possibly proton-proton, interactions to be found at energies more than  $10^7$  times higher than those at present available. □

## Negative strand viruses

from T. Barrett

A meeting on Negative Strand Viruses and the Host Cell was held at King's College, Cambridge, from 1-5 August, 1977. The proceedings will be published by Academic Press early in 1978.

NEGATIVE strand viruses are responsible for many important human and animal diseases, including influenza, Lassa fever, measles, mumps, distemper, fowl pest and rabies. This group of viruses has in common the possession of a single-stranded RNA genome which is of opposite polarity to the messenger RNA (mRNA). One of the most striking points brought out at the meeting was the great advance that has been made in the understanding of the nature of the genome of these viruses, in particular the genome of the influenza A viruses which consists of



eight separate genes. Many speakers reported methods developed to assign functions to the individual genes. P. Palese (Mount Sinai School of Medicine, New York) is mapping gene functions of influenza B strains. Unlike the A strains, the B strains do not undergo dramatic antigenic shifts and do not cause pandemics. However, like the A strains they contain eight gene segments and eight virus-specific proteins can be detected in virus-infected cells.

The genome of the bunyaviridae was discussed by D. H. L. Bishop (University of Alabama). These are arthropod-borne viruses, some of which, for example California encephalitis virus, cause disease in humans. The genome consists of three segments; two of the segments code for two virus polypeptides while the third codes for two virus glycoproteins, probably through a polycistronic message.

The segmented nature of the influenza virus genome makes it an ideal candidate for sequencing studies because individual genes can readily be isolated. P. Fellner's group (Searle, High Wycombe) in collaboration with groups in Cambridge and Mill Hill, is sequencing the two smallest genes, 7 and 8. Both have a characteristic U<sub>6</sub>A sequence near the 5' end of the viral RNA (vRNA). In prokaryotes a similar sequence is associated with transcription termination (Rosenberg *et al. Proc. natn. Acad. Sci. U.S.A.* **73**, 717; 1976) but in this case at the 3' end of the transcribed RNA rather than at the 5' end of the template. It seems that the polyadenylated messenger complementary RNA (cRNA) of influenza virus is not a full length transcript of the vRNA and the extra sequences at the 5' end may be important as recognition sites for the viral replicase. In contrast, unpolyadenylated cRNA is a full length copy of the vRNA and is presumably the template for new vRNA synthesis (A. Hay, National Institute of Medical Research, London, and R. M. Krug, Memorial Sloan-Kettering Cancer Center, New York).

Negative strand viruses with unsegmented genomes seem to have only one promoter site, and the positions of the genes in relation to the promoter site have been determined in some cases using the ultraviolet transcriptional mapping technique. The sequences of the ribosome-binding sites of all the mRNAs of vesicular stomatitis virus (VSV) have been determined by J. K. Rose (Massachusetts Institute of Technology). All contain a single AUG initiation codon. The cap structure for the N, NS and L protein mRNAs was bound by ribosomes in the initiation complex and showed a 9-nucleotide homology in the first 11 nucleotides extending from the 5' end. In contrast, the binding sites of the M

and G protein mRNAs differed both from each other and from those of the other mRNAs and did not include the cap structure. This observation has an interesting physiological aspect because translation of the M and G proteins is more sensitive to hypertonic shock (Nuss & Koch *J. Virol.* **17**, 283; 1976). A possible mechanism for this suggested by Rose is that, in the case of these two mRNAs, the ribosome needs to move to the binding site and this step may be sensitive to salt.

### Transcription and translation

Genome transcription and translation was a major topic of the meeting. A. K. Banerjee (Roche Institute, Nutley, New Jersey) proposed that synthesis of a leader RNA sequence before any mRNA synthesis, as described for VSV and Newcastle disease virus (NDV), may be a general strategy for the initiation of RNA synthesis by negative strand viruses with unsegmented genomes. All leader RNAs studied are approximately the same size and their compositions are similar. However, hybridisation studies show that there is virtually no sequence homology. The 3' end of the VSV genome has a high U content and more than half the leader sequence is composed of A residues. This probably functions as a recognition site for the virion-associated RNA polymerase. These virus genomes can be transcribed efficiently *in vitro* using a virus-associated polymerase.

A. J. Hay, T. Barrett (University of Cambridge) and J. M. Taylor (Institute for Cancer Research, Philadelphia) all found evidence for transcriptional control of individual cRNA species during influenza virus replication. Hay found that the amount of unpolyadenylated cRNAs, presumably the templates for vRNA synthesis, were produced in roughly equal amounts for each segment while the polyadenylated cRNAs, presumably the mRNAs, were produced in different amounts for each segment. Barrett and Taylor both described the preparation of cDNA copies of influenza virus RNA made using reverse transcriptase, which are being used to study the time course of vRNA synthesis.

J. Content (Institut Pasteur, Brussels) described a coupled cell-free transcription/translation system for influenza virus capable of producing authentic virus proteins *in vitro*. The proportion of viral polypeptides obtained *in vitro* was consistent with translation of primary RNA transcripts.

L. A. Ball (University of Connecticut) reported that six virus-specific gene products of NDV were faithfully expressed when both transcription and translation were carried out *in vitro*.

Ball also used the ultraviolet transcription mapping technique to determine the number of promoter sites and the gene order of NDV. The results were consistent with a single promoter site and a gene order of 3'-NP-F<sub>0</sub>-M-HN-L.

The assignment of virus cores to the conflicting demands of virus assembly and RNA synthesis was the subject of a stimulating talk by D. W. Kingsbury (St Jude Children's Hospital, Memphis). He has isolated two metabolically distinct virus core populations, free and bound, in Sendai virus. The bound fraction, inactive in RNA synthesis, contains the newly assembled cores and is likely to be the source of both the cores that enter the virion and those that become active in RNA synthesis. The free cores, active in RNA synthesis are an older population. Kingsbury suggested that a maturation process whereby more P and L protein is added to the core determines its assignment to the free RNA synthesising fraction and that once a core engages in RNA synthesis it is committed to remain within the cell and cannot be encapsidated.

H. Lodish (Massachusetts Institute of Technology) reported on the glycosylation of the virus glycoproteins at different stages in their journey from the rough endoplasmic reticulum to the cell membrane, from which they are finally incorporated into the virion. The unglycosylated form (G<sub>0</sub>) could only be detected by translation *in vitro*. A partially glycosylated form (G<sub>1</sub>), lacking the terminal sialic acid, could be detected on the rough endoplasmic reticulum, while the fully glycosylated form (G<sub>2</sub>) was detected on the smooth endoplasmic reticulum, the plasma membrane and in the mature virion.

### Virulence

One of the most interesting sessions of the meeting was concerned with aspects of virulence. Although this is a very active area of research, no clear picture of the molecular mechanism underlying virulence has emerged. H. D. Klenk (Institut für Virologie, Gießen) reported that NDV strains differ with respect to cleavage of their glycoproteins and these differences are an important factor underlying variations in NDV pathogenicity. With virulent strains of NDV, cleavage and production of infectious virus occurs in all host cells analysed, while with avirulent strains this is the case only in a few host systems. However, with influenza virus, although cleavage of the glycoproteins is indispensable for the formation of infectious virus particles, other factors also influence virus virulence. R. Rott (Giessen) suggested that a gene constellation determines virulence in the influenza virus because, as a rule,



infection with recombinants with multiple gene exchanges produces less severe clinical illness. A multigeneic determination of virulence was also favoured by J. L. Schulman (Mount Sinai School of Medicine, New York) who found that in random recombinants of a low yielding and a high yielding strain of influenza virus the genes for proteins NP, M and NS were selected in recombinants which replicated in high titres while no such selection for P genes was observed. W. J. Bean (St Jude Children's Hospital, Memphis) found that virulent recombinants always contained the NP gene from the virulent strain and, in contrast to Schulman's findings, gene 3 which codes for one of the P proteins. J. Oxford (National Institute for Biological Standards, London) isolated a clone derived from recombinants of virulent and avirulent human influenza strains which contained all the genes, except the two coding for the surface glycoproteins, from the avirulent parent but which was virulent for man. It seems that some fine coordination exists between the genes for the internal proteins of influenza virus whose balance can be upset either by mutation or recombination. This set of genes is in turn dependent on the two surface antigens for its expression.

#### Defective particles

Defective interfering virus particles are aberrant virus particles which lack parts of the virus genome and are produced when virus is passaged at high multiplicities. These DI particles interfere with the replication of normal virus particles and could be a host defence mechanism. That a host function is required for DI particle formation was shown by C. Y. Kang (University of Texas, Dallas) who found that VSV DI particles could not be produced in cells pretreated with actinomycin D, while if present in the inoculum DI particles can be replicated in the presence of the drug. DI particles can be derived from either the 5' or 3' ends of the non-defective genome. Those derived from the 5' end are capable of only a very limited RNA synthesis *in vitro* and are thought to inhibit the non-defective virus at the replicase level. J. Perrault (University of California, San Diego) and D. Kolakovsky (University of Utah, Salt Lake City) described DI particles derived from the 5' ends of VSV and Sendai virus, respectively, with inverted complementary terminal sequences. These inverted complementary sequences at the 5' end presumably confer a replicative advantage on the DI particles, since synthesis of both plus and minus strands could be achieved by the same replicase which normally recognises the

sequences at the 3' ends of the minus and plus strands.

R. A. Lazzarini (National Institutes of Health, Bethesda) described another type of DI particle of VSV derived from the 3' end of the genome. This DI particle has a leader RNA sequence and the genome is transcribed into 12–17S mRNA *in vitro* and *in vivo* without a helper virus. It lacks the greater part of the L-protein gene which is nearest the 5' end of the non-defective genome. In mixed infections, in addition to interfering with the replication of the infectious virus, it can augment the production of virus proteins, a phenomenon which may in some way trigger an enhanced immune response by the host. Clearly DI particle formation and its mode of interference is a complex process and one which will become increasingly important from the point of view of their possible role in the treatment of virus diseases. □

## Nuclear radii and magnetic moments by laser spectroscopy

from P. E. Hodgson

THE charge and matter distributions of nuclei, and their magnetic moments, are important features of nuclei, and many efforts are being made to determine them with increasingly high precision. It is often instructive to see how they change through a series of isotopes of the same element, but such studies are limited by the availability of the separated isotopes in sufficient quantities for accurate measurement. This makes it desirable to develop methods that are applicable to very small quantities of the isotope, so that measurements can be carried out on rare and on radioactive isotopes.

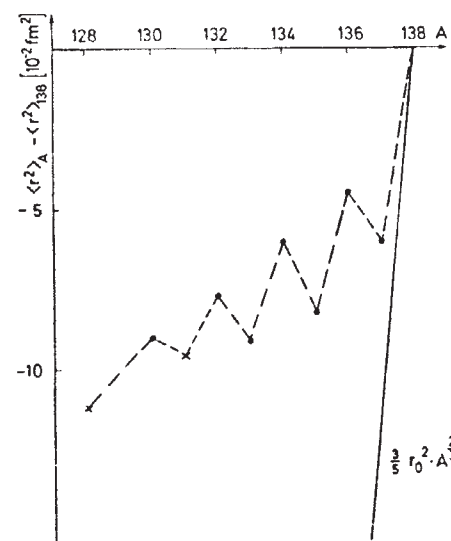
One of these methods uses high resolution laser spectroscopy, and the results of a determination of the nuclear radii and magnetic moments of  $^{128}\text{Ba}$  and  $^{131}\text{Ba}$  in this way has recently been reported by Nowicki and colleagues from the Nuclear Research Centre at Karlsruhe (*Phys. Rev. Lett.* **39**, 332; 1977).

The isotopes of barium were produced by (d,xn) reactions and subsequent  $\beta$ -decay from enriched  $^{130}\text{Ba}$  and  $^{134}\text{Ba}$ . The isotope  $^{131}\text{Ba}$  was also produced by neutron irradiation of enriched  $^{130}\text{Ba}$  in a nuclear reactor. The isotopes were enriched by an electromagnetic mass separator. This produced between 1 and 10 ng of the

isotope, sufficient for the laser method. The sample was in each case placed in an atomic-beam oven to produce a collimated beam of atoms. There is a second reference beam using the natural isotopic mixture. Light from dye lasers is passed through the atomic beams and this enables the hyperfine structure of a selected transition to be determined.

In this work the hyperfine structure of the 553.6 nm resonance transition  $6s^2\ ^1S_0$ – $6s6p\ ^1P_1$  was measured by resonance fluorescence. The frequency of this transition, and indeed of all transitions, is slightly affected by the spatial extent of the nuclear charge distribution, and a quantum mechanical analysis shows that the shift in the frequency is proportional to the mean square radius of the charge distribution. The change in frequency from one isotope to another is thus proportional to the difference between their mean square radii.

This experiment gave for the mean square radius difference between barium 131 and 130 a value of  $-0.006\text{ fm}^2$  and between barium 130 and 128 a value of  $0.022\text{ fm}^2$ . These results are shown in the figure, together with corresponding values for other isotopes. This shows that the general trend of the charge radii continues smoothly for the even isotopes and that the large odd-even staggering is markedly reduced at neutron number 75. The magnetic moment can also be calculated from the hyperfine splitting, and a value of  $-0.714\ \mu_N$  is found for  $^{131}\text{Ba}$ . □



Differences of mean square radii of the barium isotopes as obtained from optical isotope shifts. The straight line shows the dependence to be expected from a standard homogeneous sphere.

# review article

## Study of membrane permeability changes by fluctuation analysis

Charles F. Stevens\*

*The molecular basis of the changes in membrane permeability that underlie the nerve impulse can be studied by statistical analysis of the random fluctuations in current through the nerve membrane. Such "noise analysis" has already begun to clarify the nature of the gating mechanisms that control the flow of ions across the membrane at the neuromuscular junction.*

TWENTY-FIVE years ago this year Hodgkin and Huxley published a series of papers that have served as the basis for virtually all succeeding studies on the mechanism of the nerve impulse<sup>1,2</sup>. The key idea in their analysis is that the permeability of the nerve membrane to certain specific ions, such as Na and K, is controlled by the voltage difference across the membrane. Cells expend metabolic energy to keep the concentrations of the various ions within the cell different from the concentrations in the outside bathing medium—for example, the intracellular concentration of Na is about an order of magnitude lower than it is in the extracellular fluid bathing the cell—and Hodgkin and Huxley showed that the electrical signs of the nerve impulse are accounted for by the permeability changes to these ions, and the resultant ionic rush toward equilibrium. To understand the molecular basis for nervous activity, then, we must penetrate the mechanism by which cell membranes change their permeability to ions.

Despite considerable progress in our understanding of neuronal mechanisms in the past 25 years, we still do not know precisely how, at the molecular level, membranes can alter their ionic permeabilities. Quite recently—during the past five years—several new approaches to this problem have been developed, and the expectation of their success has given rise to an optimism in the field that we may soon discover the molecular basis for the nerve impulse. My goal in this review is to describe one such approach, fluctuation analysis.

The idea behind fluctuation analysis is to exploit the inherently probabilistic nature of the processes that underlie the membrane permeability changes in order to gain information about the molecular mechanisms responsible. The functioning of excitable membranes depends on the opening and closing of channels. Because these channels behave probabilistically, the exact number of open channels fluctuates around the average value. For example, if a membrane contains 1,000 independent channels, each of which has a probability of being open of 0.5, the average number of open channels will be 500. A particular membrane might actually have 490 open channels at one time, 514 at a later time, and 502 at a still later time. These fluctuations in the number of open channels are governed by the same physical processes responsible for all channel behaviour, so the study of the statistical structure of the fluctuations can give information about the underlying biophysical mechanisms. It turns out, in fact, that fluctuations can sometimes contain information that is inaccessible by other practical techniques. I shall try here to explain, in largely non-mathematical terms, how fluctuation analysis has been used to increase our understanding of membrane per-

meability mechanisms, and shall indicate how the technique can contribute in future studies. Additional information on this subject can be found in a non-mathematical survey<sup>3</sup>, and in several more technical review articles that have appeared recently<sup>4-6</sup>.

My explanations about fluctuation analysis will be chiefly in the form of specific examples, first drawn from investigations of synaptic channels that are caused to open by the neurotransmitter acetylcholine, and second from the nerve membrane itself. Before giving these examples, however, I shall review briefly current views about the molecular basis for the nerve impulse, and survey some standard techniques for characterising fluctuations.

### Channels and gates

Changes in permeability to ions are responsible for the nerve impulse. How do these changes occur? The current view is that nerve cells have intrinsic membrane proteins that control the permeability to specific ions by providing channels through which they can pass. Each channel type is named for the ion that normally best passes through it, so one speaks of "sodium channels" and "potassium channels", even though potassium ions can move, for instance, through "sodium channels" to some extent.

Understanding how these membrane proteins regulate permeability to ions depends on the answers to two central questions. The first is: what are the molecular mechanisms responsible for the selective permeability of channels? For example, what is the physical basis for an ion-channel interaction that permits potassium ions (crystal radius = 1.33 Å) to pass through and nearly excludes sodium ions (crystal radius = 0.99 Å). The second central question relates to mechanisms by which membrane proteins alter ionic permeability. How, for example, is a membrane able to increase its permeability to sodium ions by many orders of magnitude in less than a millisecond? Questions of this second sort are said to deal with channel "gating", because increases in permeability occur as if gates obstructing the flow of ions through channels suddenly opened. This review is concerned primarily with gating rather than with selectivity.

Over the years since Hodgkin and Huxley first studied gating of sodium and potassium channels in the squid giant axon considerable information about this process has accumulated. There is now general agreement that gating occurs by conformational changes in a macromolecule. These conformational changes are driven principally by interactions of the gate with the membrane electric field (membrane field strengths change by about 100 kV cm<sup>-1</sup> during the nerve impulse) in electrically excitable membranes such as those of nerve cells, or by the binding of a ligand known as a neurotransmitter in chemically activated membranes such as those that occur at synapses. Ultimate

\*Department of Physiology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510



understanding of gating would then involve knowing how electric fields or ligands can induce conformational transitions, knowing the number of possible conformational states the gate can assume and the rates of transition between these states, and finally relating this information to the structure of the gating macromolecule.

One of the main approaches to understanding the behaviour of gates is through the study of kinetics. This approach was instituted by Hodgkin and Huxley and involves measuring the rates at which channels go from the open to the closed state in various circumstances. For example, one might drive channels from the closed to the open state by causing a step-like change in the voltage across the nerve membrane. The time course of opening would be monitored by measuring the permeability of the membrane to, say, sodium ions as a function of time after the voltage step. One then tries to make inferences about the nature of the underlying conformational changes from the kinetics of channel opening revealed in such an experiment. Usually, experiments measure only the average number of channels open as a function of time in the sort of experiment just described, but more recently membrane biophysicists have been studying the fluctuations around the mean number of open channels. Such fluctuations occur because channel opening and closing is a probabilistic process. Although the investigation of these fluctuations is merely a variation on the more usual kinetic analysis, additional information can sometimes be gained from the fluctuations. My purpose here is to indicate why this is so.

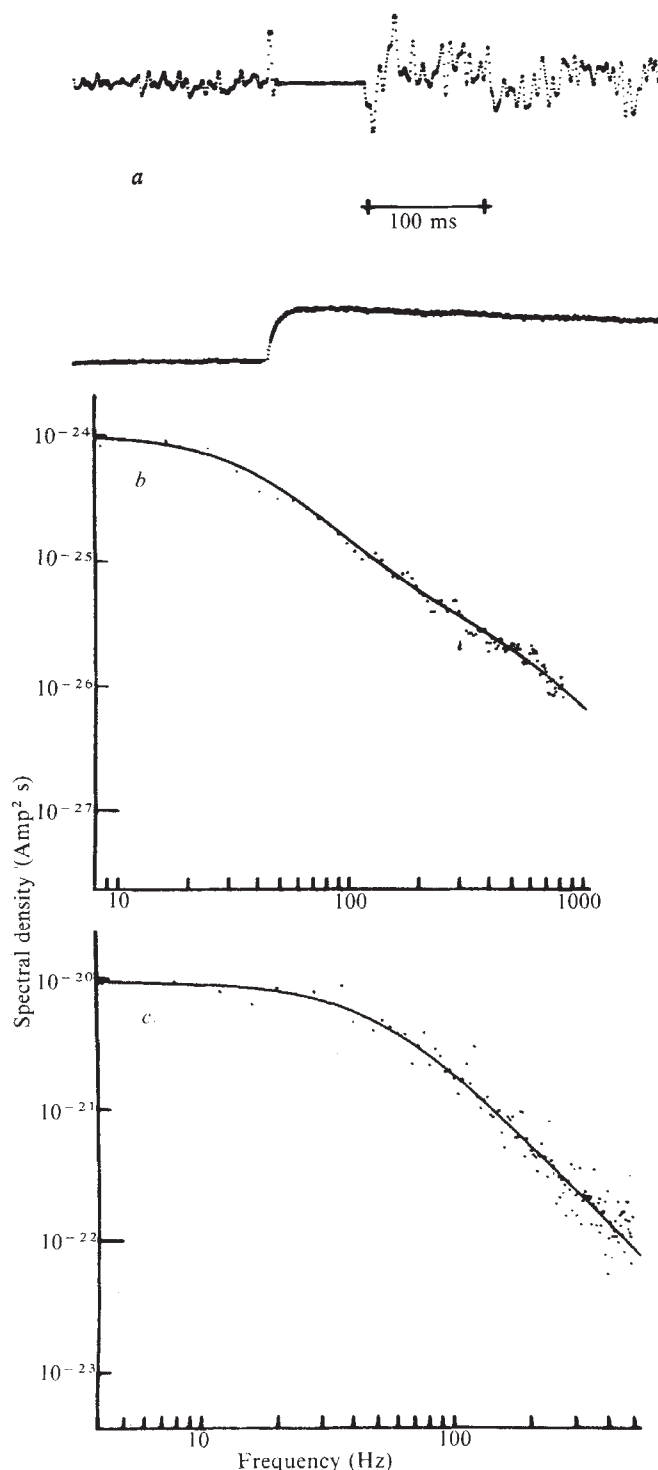
### The structure of noise

Records of a fluctuating quantity, such as the number of open channels, have a statistical structure, and information about a system can be inferred from the character of this structure. Before giving examples of uses of fluctuation analysis, then, I shall summarise briefly some standard characterisations of noise. The following discussion will usually assume that the systems under study are stationary. This means that the statistical properties of the fluctuations do not change with time over the course of a set of observations. For example, if the fluctuating quantity under study is the number of open channels, the exact number of channels open varies from instant to instant, but the average number open must stay fixed for the length of the sample; if the average number of channels open is  $1/2$  at the start of the observation period, an average of  $1/2$  must be open throughout the period.

The simplest characterisation of fluctuations is the variance, that is, the average of the squared deviations from the mean. The variance is a usual measure for the amplitude of fluctuations.

The variance indicates the overall magnitude of fluctuations, but gives no information about how rapidly they vary with time. Random signals can, however, differ dramatically in how fast the wiggles occur with, for example, one type of system producing a signal dominated by low frequency fluctuations and another by higher frequencies. Much of the information about a system that can be extracted from its fluctuations is contained in this frequency structure. A natural way to extend the simple characterisation by variance to give information about the rapidity of fluctuations is to find the contribution that each frequency component of a random signal makes to the overall variance. A record of fluctuations could be decomposed by Fourier analysis into its various frequency components (for example, 1 Hz, 2 Hz, 3 Hz, . . . . ., 100 Hz), and the variance per Hertz contributed by each component could be computed. If the random signal generated by the system were slowly varying, the variance contributed by the higher frequencies would be small, whereas a rapidly changing signal would have a much larger contribution from the high frequency components. The collection of numbers that specify the variance per Hertz for the different frequency components of a random signal is known as the spectrum associated with that signal. Figure 1 presents examples of two experimentally determined spectra displayed in the usual double logarithmic plots.

The individual record of fluctuations produced by a probabilistic system is almost never exactly the same twice, but the



**Fig. 1** Sample fluctuations, and spectra from nerve and synaptic membranes. *a*, Fluctuations in potassium current and mean potassium current (proportional to potassium conductance) from a voltage-clamped frog node of Ranvier (T. Begenisich and C.F.S., unpublished observations). Lower trace is mean current through potassium channels as a function of time, with the potassium conductance activated by a positive going voltage step. Upper trace is simultaneous recording at  $40\times$  greater gain with the mean current subtracted out. Fluctuations on the left of the upper trace represent equipment noise contributions. Fluctuations due to channel opening and closing are visible on right of trace after the period of flat base line which represents a blanking pulse to suppress filter transients. *b*, Spectrum of frog node of Ranvier potassium current fluctuations such as those in *a* (from T. Begenisich and C.F.S., unpublished observations). *c*, Spectrum of fluctuations in current through acetylcholine activated channels at the frog nerve-muscle junction (C. Lewis and C.F.S., unpublished observations). Acetylcholine was iontophoretically applied to a voltage clamped post-junctional membrane, and the resulting current fluctuations (proportional to conductance fluctuations) Fourier analysed to give the spectral densities as a function of frequency plotted above.

spectrum characterising the fluctuations is, except for sampling errors, always just the same. This spectrum reflects the characteristic times of the underlying molecular processes, and in fact the reason for determining a spectrum is that it gives a way of measuring these characteristic times. If underlying processes are fast, the spectrum will reveal large contributions from high frequencies, and if the system is sluggish, only low frequencies will be appreciable.

Probability theory provides standard recipes for calculating what spectrum any underlying probabilistic mechanism will produce<sup>7,8</sup>. If one can formulate a probabilistic theory for a process, that theory provides predictions not only about the system's average behaviour, but also about the spectrum associated with the fluctuations about this average. Comparing predicted spectra with observed ones thus provides a way of testing for the accuracy of some proposed mechanism. It is extremely important to emphasise that no automatic way exists for using a spectrum to gain information about a system. The study of fluctuations can, as will be explained later, give information that might otherwise be inaccessible for practical reasons, but all inferences depend entirely on some sort of theory about the source of the fluctuations.

### Synapse channels

Fluctuation analysis has been most extensively used to study the permeability changes produced by the application of the neurotransmitter acetylcholine to postsynaptic membranes, and this system will thus be used for the first examples. Although the examples presented here do not by any means include all of the applications of this technique, they do illustrate types of use. Application of fluctuation analysis to postsynaptic membranes has recently been reviewed<sup>6</sup>.

When a nerve impulse reaches a synaptic nerve terminal on a muscle it causes the release of acetylcholine from the nerve ending. The acetylcholine diffuses to a specialised postjunctional region of the muscle membrane where it causes a transient permeability increase to sodium and some other ions. The permeability to these ions increases very rapidly and then decreases again over a several millisecond period. These permeability changes are the first step in a sequence that leads to muscle contraction, and they can be measured by recording the currents that flow through the muscle membrane. These currents that reflect the permeability increase caused by acetylcholine are called endplate currents.

Until fairly recently, workers in the field had generally believed that the endplate current was a fairly faithful representation of the time course of acetylcholine concentration at the postjunctional membrane. Acetylcholine is removed from the vicinity of the postjunctional membrane by diffusion and enzyme action, and the assumption had been that it took several milliseconds for the acetylcholine to disappear by these processes and that endplate current continued to flow for as long as the acetylcholine was present. An analysis of the acetylcholine-produced permeability increase at the frog neuromuscular junction led Magleby and Stevens<sup>9</sup> to propose a theory that supposed the rate limiting step in channel closing is a conformational change in the gating molecule rather than a decline in acetylcholine concentration. According to this theory, one must assume that acetylcholine remains near the postjunctional membrane only for a very short time. During the brief action of acetylcholine, channels open and then close probabilistically over a several-millisecond period after the acetylcholine has been removed. This theory accounts quantitatively for many properties of the endplate current, but the assumption that the channel-closing conformational change is rate limiting could not be put to direct test with the techniques available. The difficulty was that the acetylcholine concentration at the postjunctional membrane could not be rapidly measured or controlled. This difficulty was circumvented by using fluctuation analysis.

Katz and Miledi had discovered that sizeable fluctuations in muscle membrane permeability occur when a constant concentration of acetylcholine is present at the neuromuscular

junction, and they concluded that these fluctuations represent the random opening and closing of channels<sup>10</sup>. The scheme Magleby and Stevens developed to describe how acetylcholine causes permeability increases, being inherently a probabilistic theory, predicted the spectrum of these fluctuations. Specifically, channels were assumed to have just two conductance states, open and closed, and a channel was viewed as fluctuating randomly between these two states when acetylcholine was bound to a specific receptor associated with the channel. This theory predicts<sup>9</sup> that the spectral density  $S(f)$  (contribution to the total variance) of the component of fluctuations with frequency  $f$  should be given by

$$S(f) = \frac{4\gamma g/\alpha}{1 + (f/2\pi\alpha)^2} \quad (1)$$

where  $g$  is the average conductance increase caused by acetylcholine,  $\alpha$  is the average rate at which open channels make the transition to the closed state, and  $\gamma$  is the conductance of a single open channel. The average conductance can be measured in the same experiment that produces records of fluctuations, and  $\alpha$  can be obtained through the Magleby-Stevens theory, independently of the fluctuation measurements, from endplate currents produced by nerve stimulation. The only parameter in this equation not found independently was the single channel conductance  $\gamma$ , and this was known to have an approximately constant value under a variety of experimental conditions that produced quite large changes in  $\alpha$  and  $g$ .

The strategy then is to observe the conductance (that is, effectively permeability) fluctuations and calculate their spectrum. If the theory is correct, equation (1) should describe the spectrum accurately with only one constant  $\gamma$  to be estimated from the fluctuations. As various conditions are changed that alter the spectra,  $\gamma$  should remain constant. If the channel closing step is not rate limiting, or if the theory is incorrect in a number of other ways, equation (1) should not accurately characterise the experimentally determined spectra under the various different experimental conditions.

Using this approach, Anderson and Stevens<sup>11</sup> found that equation (1) was indeed an accurate description of the spectra. This success of the theory gave strong support to the idea that channel gating properties, rather than loss of acetylcholine from the nerve-muscle junction region, are rate limiting during the normal endplate current. Furthermore, once the theoretical interpretation of the source for fluctuations could be accepted, these same experiments gave a measurement of  $\gamma$ , the single channel conductance.

Two properties of individual channels may thus be estimated from fluctuations in permeability through the use of equation (1), the single channel conductance  $\gamma$  and the average closing rate for open channels  $\alpha$ . The statistical theory that led to equation (1) provides some additional information about the channel behaviour: the average length of time  $\tau$  that a channel, once open, remains open is just  $1/\alpha$ . Fluctuation analysis therefore gives a way of measuring the conductance of a channel, and the mean length of time it stays open without ever actually making measurements on a single channel.

Generally various inevitable sources of instrumentation noise make it impossible to record directly, and in isolation, single channel currents. By carefully analysing the sources of noise in current measurements and cleverly designing apparatus that minimises these noise contributions, however, Neher and Sakmann<sup>12</sup> have recently been able to record the currents flowing through individual acetylcholine-activated channels in frog postjunctional membrane. Their results are in accord with the picture developed on the basis of fluctuation analysis: channels seem to have only two conductance states (open and closed), and the duration of the open state is exponentially distributed. The mean open time and  $\gamma$  have values in agreement with those determined by investigation of fluctuations arising from many channels independently opening and closing. Neher's and Sakmann's achievement is particularly important because cer-



tain assumptions made in the theory on which fluctuation analysis is based could not be independently checked. For example, all of the data obtained before their direct demonstration of simple open-closed behaviour were equally consistent with the idea that a channel opened instantaneously and then decreased its conductance exponentially with time.

Fluctuation analysis has been used to answer a number of questions about acetylcholine-activated channels<sup>6</sup>. In some instances, the quantities of interest were single channel conductance or mean open time and equation (1) had to be used to estimate these quantities because they could not be found by alternative means. In other instances certain experimental manipulations led to more complicated situations in which the spectra had a more complex form than equation (1) and an equation for the spectral density had to be derived from an analysis of mechanisms. I will give two specific examples that illustrate these approaches.

Colquhoun *et al.*<sup>13</sup> have recently used fluctuation analysis as an aid in studying acetylcholine metabolism at the rat nerve-muscle junction. Their approach depends upon the important discovery by Katz and Miledi<sup>14</sup> that various analogues of the natural neurotransmitter acetylcholine each cause the channel to remain open for a different mean time for a given set of conditions. For example, the mean open time is shorter for carbacol than for acetylcholine, and longer for suberyldicholine than for acetylcholine. Colquhoun *et al.* found conditions under which monoethylcholine would be taken up by the nerve terminal and synthesised into acetylmonoethylcholine. This "false" neurotransmitter was then released by the nerve on stimulation. The fact that acetylmonoethylcholine was being released could be detected because channels stayed open a shorter average time with this false transmitter than with the normal acetylcholine, and the false transmitter was less potent than the normal one. The characteristic channel behaviour elicited by the false transmitter was found by using fluctuation analysis to measure channel mean open time and single channel conductance for externally applied acetylmonoethylcholine. Being able to detect the quantity of false and natural transmitter released means one can study the details of vesicle filling with more precision than otherwise possible.

Ruff<sup>15</sup> has recently used fluctuation analysis to help unravel the action of local anaesthetics at the postjunctional membrane. Normally the endplate current decays according to a simple exponential time course (reflecting the conformational change in closing) but, after treatment with local anaesthetics, the decaying phase of the endplate current exhibits two main exponentials, one faster and one slower than normal. Of the various hypotheses that could account quantitatively for this behaviour, two specific ones were: (1) the local anaesthetic molecules alter the lipid microenvironment of receptors and cause islands of lipid and proteins with two different characteristics. Some islands are stiffer than normal, whereas others are more fluid. In the more fluid islands, the conformational change can occur more rapidly (this gives the rapid exponential decay), whereas the stiffer environment produces slower than normal conformational changes (this gives the slower exponential component in the decay of endplate current). (2) Local anaesthetic molecules could rush in and block open channels so that ion transport through the membrane could not proceed but nor could the conformational change occur that closes the gate. In this theory the rapid exponential in the decay would reflect the initial establishment of a steady state number of blocked channels, and the second exponential would arise from the slow unblocking and closing step.

Both of these theories, as well as others, could account quantitatively for the decay of endplate currents and no observation on mean endplate currents can distinguish between them because they give rise to the same descriptive equation. The theories do, however, make different predictions about the spectra. Both mechanisms would give a spectrum with two components that is more complicated than the type described by equation (1). The amplitudes of the components are different for the two theories, so that a measurement of the mean endplate current and the spectrum permits a differential test. Ruff has carried out this procedure for the two theories indicated above

and others, and has found that the blocking mechanism is in quantitative agreement with the experimental observations, but predictions of alternative theories are not. This example illustrates a fairly common occurrence: different molecular mechanisms make identical or practically indistinguishable predictions about average behaviour, but give rise to spectra with easily detectable differences.

## Nerve channels

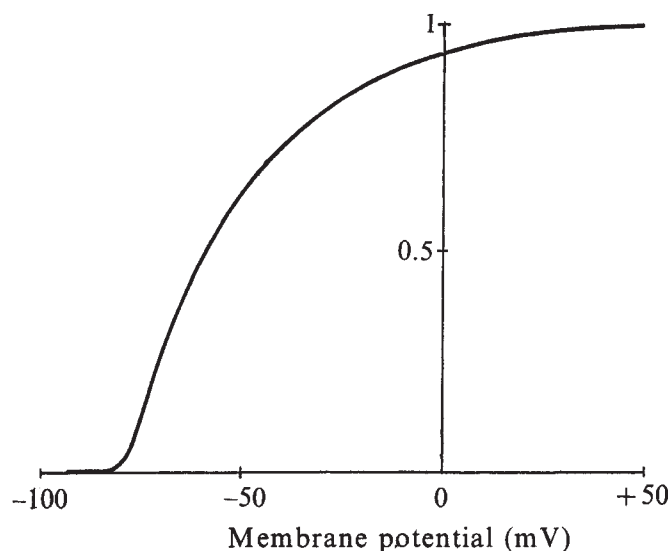
In 1966 Verveen and Derksen published a short account of their studies on fluctuations in nerve membrane permeability<sup>16</sup> and thus began the use of fluctuation analysis to investigate nerve membrane properties. Interestingly, the main motivation for the early studies from Verveen's laboratory related more to information processing properties of the neurones than to the molecular mechanisms of excitability. Verveen's thesis<sup>17</sup> had been concerned with random threshold fluctuations in nerve, and he and Derksen had set out to discover the membrane correlates of this phenomenon. Blair and Erlanger<sup>18</sup> had, in the course of their pioneering work on the nerve impulse, discovered that nerves stimulated near threshold produce nerve impulses probabilistically, and Pecher<sup>19</sup> had quite carefully investigated this phenomenon and found that the threshold variations were independent and normally distributed if stimuli were applied at an appropriate rate. Early interest thus centred on the probabilistic behaviour neurones exhibited in some circumstances, and was largely motivated by implications this behaviour has for nervous system function: how can the nervous system be reliable when its basic elements, the neurones, are "unreliable" in the sense that they cannot be counted on to give a nerve impulse whenever stimulated?

In recent years, however, the main motivation for the study of fluctuations in properties of the nerve membrane has been related not to information processing, but rather to the significance for the molecular basis of nerve excitability. Furthermore primary interest has shifted from studies of the so called "1/f noise" first described by Verveen and Derksen to the conductance fluctuations that arise from the probabilistic opening and closing of channels. In the remainder of this review I shall concentrate on "channel noise" rather than on the 1/f noise associated with ion motion through open channels; that is, my attention will be limited to fluctuations in conductance associated with the opening and closing of the channels responsible for nerve excitability.

A resting nerve cell normally has an interior that is about 60 mV more negative than the outside solution. If the voltage across the cell membrane is displaced from the normal resting value of -60 mV to more positive values, ion selective channels (Na, K, Ca) are caused to open so that the membrane conductance increases. Hodgkin and Huxley discovered that the increase in conductance as a function of voltage is a sigmoid function of voltage for each channel type (Fig. 2).

Two limiting mechanisms can be envisaged for this type of behaviour. On the one hand, one could suppose that each channel opened in a graded manner with voltage. According to this view, we might expect that, if the conductance of just one channel could be measured at different voltages, a single channel would show a conductance against voltage plot like that exhibited in Fig. 2. On the other hand, each channel might have only two conductance states, open and closed, and the gross conductance against voltage curve (Fig. 2) might simply reflect the fraction of channels open at any voltage. Of course, intermediate cases are easily envisaged. A channel might have, say, four different conductance levels (closed, 1/3 open, 2/3 open, fully open), and Fig. 2 would reflect the fraction of channels in each of their conductance states. Clearly, one cannot formulate a theory for channel opening mechanisms without knowing which of these possibilities is correct. Fluctuation analysis has been used to investigate this important question, and is, unfortunately, the only approach currently available.

At the acetylcholine-activated channels of the nerve-muscle junction increasing the dose of acetylcholine causes a larger conductance by causing more channels to open, but  $\gamma$  has the



**Fig. 2** A typical plot of steady-state conductance for one channel type (potassium, for example) as a function of voltage difference between the interior of a nerve cell and the exterior solution. The ordinate presents conductance as a fraction of the maximum possible value.

same value irrespective of how many channels are opened on the average. For the nerve membrane, channels open in response to the membrane voltage rather than the presence of a neurotransmitter like acetylcholine, but in this case as well,  $\gamma$  should be independent of the number of open channels if each channel has only two conductance states, open and closed. If, on the other hand, each channel has many conductance states, with the higher conductances favoured by positive voltages, one might expect that  $\gamma$  would increase as the total conductance were increased by subjecting the membrane to positive voltages. Although the actual situation turns out to be somewhat more complicated than these simplest expectations, fluctuation analysis can show whether  $\gamma$  remains constant as the total conductance increases.

The use of fluctuation analysis to get an indication of how many conductance states a type of channel might have is best introduced by considering the limiting case in which channels can be only open and closed. According to simple coin-flipping theory<sup>7,8</sup>, we know that the average conductance is

$$g = \gamma N f \quad (2)$$

where  $\gamma$  is the single channel conductance,  $g$  is the average total membrane conductance,  $N$  is the number of channels in the membrane (they are assumed to be independent), and  $f$  is the probability that a channel is open. The quantity  $f$  would depend on voltage, and thus so would  $g$ . The variance in conductance is

$$s^2 = \gamma^2 N f (1-f); \quad (3)$$

where  $s^2$  is the variance and other symbols have the same meaning as in equation (2). The quantities  $g$  and  $s^2$  can be measured in experiment, and  $f$  can be estimated from the conductance/voltage function such as appears in Fig. 2 if we assume that maximum conductance corresponds to a channel certainly being open ( $f$  is then just the fraction of the maximum conductance). Eliminating the unknown quantity  $N$  from equations (2) and (3), we can solve for  $\gamma$ , the single channel conductance, in terms of measurable quantities  $g$ ,  $s^2$ , and  $f$ :

$$\gamma = \frac{s^2}{g(1-f)} \quad (4)$$

Equation (4) gives a means of estimating single channel conductance if channels have only two conductance states.

If the excitable membrane channels have only two conductance states then  $\gamma$  estimated by equation (4) should, like the acetylcholine-activated channels described earlier, always have the same value. (Actually channels show rectification under some circumstances, so  $\gamma$  might appear non-constant. This rectification can be taken into account by measuring the instantaneous voltage-current relation and using the information to refer values of  $\gamma$  to some reference voltage.) As the membrane voltage is made more positive the probability that a channel is open approaches unity; the average number of open channels thus increases with positive voltages, so the average conductance  $g$  might increase along a sigmoid curve like that illustrated in Fig. 2. The variance in conductance fluctuations should, however, follow an inverted U-shaped function of voltage (according to equation (3)) that gives a constant  $\gamma$  estimate.

What would be expected if single channels have multiple conductance states so that the average conductance of a channel increases as the membrane voltage is made more positive? In general,  $\gamma$  as defined by equation (4) would not for this case be a constant independent of membrane voltage, but the exact behaviour of  $\gamma$  as the total conductance increased would depend on the exact mechanism underlying the conductance increase. Specifically,  $\gamma$  (according to equations (4)) might increase, decrease, or remain constant as  $g$  was increased by making the membrane voltage more positive. If the primary means by which the membrane increases its overall conductance is by making each channel more conductive,  $\gamma$  would, on the simplest theories, increase as  $g$  did. It is not difficult, however, to construct specific examples in which  $\gamma$  remains constant or decreases as  $g$  is increased.

In summary, measuring  $\gamma$  as a function of the mean conductance  $g$  can give an indication about whether individual channels have two or more than two conductance states. If a channel can only be open and closed,  $\gamma$  must be constant. If  $\gamma$  calculated from equation (4) is actually observed to be constant, the simplest interpretation is that channels can be only open or closed. The experiments performed so far indicate that  $\gamma$  is constant as  $g$  increases, so gates seem to behave in the simple open-closed way.

Three investigations of single channel conductance, estimated by equation (4), as a function of the total conductance  $g$  have appeared, all on frog nerve; one study was on potassium channels<sup>20</sup>, and two on sodium channels<sup>21,22</sup>. Begenisich and Stevens<sup>20</sup> could find no significant change in  $\gamma$  for K channels over a range of voltages that give from 15% to 85% of the maximum conductance. These workers found that the single channel conductance was 4 pS ( $4 \times 10^{-12}$  ohm<sup>-1</sup>), a value about 7 times lower than that for the acetylcholine-activated channels.

Sodium channels are more difficult to investigate than are potassium channels because of sodium channel inactivation: at most voltages, sodium channels open only transiently. The opening process is called activation and the closing, in the face of a maintained positive voltage, is known as inactivation. For example, if the membrane voltage is changed from a resting value of -60 mV to +40 mV, the sodium conductance would rapidly increase, and then decrease to nearly zero even though the voltage was kept constant. After becoming inactivated, sodium channels must have a rest period at a negative voltage before they can be activated again. This process of inactivation not only complicates the theoretical analysis of sodium channels, but also hinders the application of fluctuation analysis that requires observation to be made during a relatively long period during which the average conductance remains nearly constant. Potassium channels also show a similar inactivation phenomenon, but the rate of development of the inactivated state is so slow that sufficiently long samples with an insignificant change in mean conductance can be obtained.

Conti *et al.*<sup>22</sup> have studied fluctuations to investigate the behaviour of the sodium channel  $\gamma$  as the average conductance was varied. Because of the presence of sodium channel inactivation, however, these workers were restricted to a somewhat smaller range of voltages than were Begenisich and Stevens. As



with the potassium channel,  $\gamma$  remained constant over the entire accessible voltage range with a value of about 8 pS.

More recently Sigworth<sup>22</sup> has approached the problem with a different method that does not assume stationarity and that therefore circumvents the difficulties arising from the presence of sodium channel inactivation. The studies described until now have all carried out an analysis of fluctuations that occur over time around a constant or nearly constant mean value. Sigworth collected an ensemble of records with a transient rise and fall in sodium conductance. These records should all have been identical except for probabilistic fluctuations. After evaluating the source of variability other than from the random opening and closing of channels, Sigworth calculated the variance about the mean from all records at each time point. Equation (4) should still hold, except that the mean and variance are calculated for an ensemble of records rather than over time for a single record.

Using this approach, Sigworth was able to estimate  $\gamma$  over the entire range from the resting voltage to +150 mV. In confirmation of the findings of Conti *et al.*, these experiments give a constant  $\gamma$  with a value of 8 pS under normal conditions. The value of  $\gamma$  was unchanged by treatment with tetrodotoxin, agent that decreases the number of functional channels present, but was decreased by lowering pH. This pH effect is consistent with a blocking action of hydrogen ions on ionic conduction through the channel<sup>23</sup>.

Although these studies are certainly not conclusive, they do constitute the first experimental support for the notion that channels in the nerve membrane, like the acetylcholine-activated channels at the nerve-muscle junction, have only two main conductance states, open and closed. If sustained by future research, this conclusion greatly simplifies the understanding of gating mechanisms.

The first application, of fluctuation analysis to nerve membrane, thus relates to evidence for an all-or-none gate on potassium and sodium channels. Another use, and one that will probably be more important, is in distinguishing between various possible molecular mechanisms of gating. Even at a constant voltage, the time course by which nerve channels open and close is quite complicated and requires a fairly high order differential equation for a good description. The Hodgkin-Huxley equations for potassium channels, for example, involve a fourth order differential equation. The difficulty is that various possible underlying mechanisms can give rise to the same equation for average behaviour, and an even larger number of mechanisms is consistent with similar mathematical descriptions that would, in

practice, be indistinguishable in their predictions for average behaviour. In general, however, distinct molecular mechanisms that predict identical or very similar average behaviour can make quite different predictions about the spectra of conductance fluctuations. Fluctuation analysis, then, has the potential of helping to decide between various alternative mechanisms that equally well account for the average behaviour of nerve channels. This approach has already been successfully used by Ruff, as described earlier, in investigations of local anaesthetic actions.

A number of investigators have published spectra of conductance fluctuations for nerve membranes<sup>21,24-29</sup>. So far, however, for reasons that are not yet clear, no two laboratories have managed to agree on the form and amplitudes of the spectra they find. The nature of these disagreements has recently been detailed by Nether and Stevens<sup>6</sup> and will not be repeated here.

To summarise, one may say that fluctuation analysis has already helped in understanding gating mechanisms at the nerve-muscle junction as indicated by the examples given earlier. Further the technique has given the first experimental evidence that nerve channels gates behave in all all-or-none manner. The challenge for the future will be to make use of this approach to elucidate the processes underlying the phenomena described by the Hodgkin-Huxley equations.

1. Hodgkin, A. L., Huxley, A. F. & Katz, B. *J. Physiol., Lond.* **116**, 424-448 (1952).
2. Hodgkin, A. L. & Huxley, A. F. *J. Physiol., Lond.* **116**, 449-472 (1952); *ibid.* **116**, 473-496 (1952); *ibid.* **116**, 497-506 (1952); *ibid.* **117**, 500-544 (1952).
3. Stevens, C. F. *Fedn. Proc.* **34**, 1364-1369 (1975).
4. Verveen, A. A. & DeFelice, L. J. *Prog. Biophys. molec. Biol.* **28**, 189-265 (1974).
5. Conti, F. & Wanke, E. *Q. Rev. Biophys.* **8**, 451-506 (1975).
6. Neher, E. & Stevens, C. F. *Ann. Rev. Biophys. Bioeng.* **6**, 345-381 (1977).
7. Feller, W. *An Introduction to Probability Theory and Its Applications* **2** (Wiley, New York, 1966).
8. Moran, P. A. P. *An Introduction to Probability Theory* (Clarendon, Oxford, 1968).
9. Magleby, K. L. & Stevens, C. F. *J. Physiol., Lond.* **223**, 151-171 (1972); *ibid.* **223**, 173-197 (1972).
10. Katz, B. & Miledi, R. *Nature* **226**, 962-963 (1970); **224**, 665-699 (1972).
11. Anderson, C. R. & Stevens, C. F. *J. Physiol., Lond.* **235**, 655-691 (1973).
12. Neher, E. & Sakmann, B. *Nature* **260**, 779-802 (1976).
13. Colquhoun, D., Large, W. A. & Rang, H. P. *J. Physiol., Lond.* **266**, 361-395 (1977).
14. Katz, B. & Miledi, R. *J. Physiol., Lond.* **230**, 707-717 (1973).
15. Ruff, R. L. *Biophys. J.* **16**, 433-439 (1976); *J. Physiol.* **264**, 89-124 (1977).
16. Derksen, H. E. & Verveen, A. A. *Science* **151**, 1388-1389 (1966).
17. Verveen, A. A. *Fluctuation in Excitability* 1-76 (Drukkerij Holland N.V., Amsterdam, 1961).
18. Blair, E. A. & Erlanger, J. *Am. J. Physiol.* **106**, 524-564 (1933).
19. Pecher, C. *Arch. Internat. Physiol.* **49**, 129-152 (1939).
20. Begenisich, T. & Stevens, C. F. *Biophys. J.* **15**, 843-846 (1975).
21. Conti, F., Hille, B., Neumcke, B., Nonner, W. & Stämpfli, R. *J. Physiol., Lond.* **262**, 699-727 (1976); *ibid.* **262**, 729-742 (1976).
22. Sigworth, F. J. *Nature* **270**, 265-267 (1977).
23. Woodhull, A. M. *J. gen. Physiol.* **61**, 687-708 (1973).
24. Siebenga, E., Meyer, A. W. A. & Verveen, A. A. *Pflügers Arch.* **341**, 87-96 (1973).
25. Fishman, H. M. *Proc. natn. Acad. Sci. U.S.A.* **70**, 876-879 (1973).
26. Fishman, H. M., Moore, L. E. & Poussart, D. J. *Memb. Biol.* **24**, 305-328 (1975).
27. Conti, F., DeFelice, L. J. & Wanke, E. *J. Physiol., Lond.* **248**, 45-82 (1975).
28. Fishman, H. M. *J. membr. Biol.* **24**, 265-277 (1975).
29. van den Berg, R. J., de Goede, J. & Verveen, A. A. *Pflügers Arch.* **360**, 17-23 (1975).

## articles

### Hubble's constant determined from super-luminal radio sources

D. Lynden-Bell

Institute of Astronomy, The Observatories, Cambridge, UK

*Recent data on the so-called super-light expansion velocities observed in the radio galaxy 3C120 show a good fit to the light echo theory, provided the Hubble constant is  $110 \pm 10 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . Statistics on three super-luminal sources interpreted with that theory give the same Hubble constant. A method of determining  $q_0$  from each source with data that may already exist, looks very promising.*

REES and others<sup>1,2</sup> have displayed remarkable ingenuity in devising mechanisms that can lead to high multiples of  $c$  in apparent transverse motion, nevertheless, the only theory that looks plausible for general application is the old 'light echo' theory<sup>3</sup>. In 1901 Nova Persei exploded close to a reflection nebula and its light echo was seen on the nebula with an apparent transverse expansion of a diameter of  $2c$ , and was used to determine a distance to Nova Persei<sup>4</sup>. I aim to show here that the

mystery that surrounds the super-luminal velocities observed in radio sources is a myth, and that all well observed objects can be explained with this old theory. The mystery has developed because most authors preferred too low a value of the Hubble constant and the interpretation of the data has caused some confusion. This investigation was stimulated by the account of Cohen *et al.*<sup>5</sup> of both super-luminal velocities and the ambiguities of their interpretation.

My accretion disk picture<sup>6</sup> of quasars and Seyfert galaxies<sup>7-10</sup> has the inner part of the disk inflated into a nearly spherical shape by radiation pressure<sup>10,11</sup>. Just as water flowing into the black hole in a bath does so around a narrow vortex, so the sphere is pierced by a narrow vortex hole on the sides of which velocities close to  $c$  are reached as the central black hole is approached. X rays, strong electromagnetic waves and relativistic plasma generated close to the hole can only escape up or down the narrow tube formed by this vortex. I shall assume that brilliant flashes of radiation (or alternatively relativistic shock waves moving with velocity  $c$ ) emerge up these oppositely-directed holes and that their energy density is so great that they generate relativistic random motions in the plasma they illuminate. This causes the plasma to radiate by the synchrotron process. The patches illuminated by a flash from the centre each move out with the velocity of light. If the tube were perpendicular to the line of sight the apparent transverse expansion apart of the two components would be  $2c$ . Allowance for an angle  $\theta$  between the tube and the observers's line of sight, however, changes this to  $2c/\sin \theta$ .

There are two immediate objections to this theory. (1) 3C120 has shown two flashes at  $5c$  and  $8c$  so that  $\theta$  must change over a short period, but the position angle on the sky remains unchanged. I shall show that the second expansion velocity is caused by a misinterpretation and that the events in this galaxy fit the theory perfectly. (2) There are too many very large multiples of  $c$  observed in apparent expansions. This is explained by the change of Hubble constant together with some over-concentration on more esoteric interpretations of observed events. In particular Cohen *et al.*<sup>5</sup> point out that events in 3C279 can either be interpreted as a  $10c$  expansion ( $5c$  on the new Hubble constant) or as three fixed sources of variable intensity. No doubt two moving sources and an off-central source of variability would be as good.

The light echo theory used here suggests rules of interpretation which should help to build much better models from very long baseline (VLB) data. These in turn will yield values for  $q_0$  the deceleration parameter that determines the geometry of the universe. Data probably already exist which will allow this to be done for 3C345, 3C279, 3C454.3.

Here a simple account of the light echo theory and its predictions is given. These are used to determine the Hubble constant in two different ways. First, from the data on 3C120 and second, from the statistics of the three well-observed super-luminal expansions.

### Simple light echo theory for non-cosmological distances

Consider an object O not moving with respect to the radio antenna A. Let its tube point at  $\theta$  to the line of sight. Then a flash from O will illuminate and excite material at a source  $S_1$  distant  $ct$  from it. This event will be seen by A not at time  $t$  after he sees the initial burst at O but at time  $t - (ct \cos \theta)/c$  because the source  $S_1$  is nearer than O. The observed transverse velocity of  $S_1$  is

$$[ct \sin \theta]/[t(1 - \cos \theta)].$$

That is

$$v_1/c = \sin \theta/(1 - \cos \theta) \quad (1)$$

The source flying in the opposite direction has apparent transverse motion  $v_2$  where

$$v_2/c = \sin \theta/(1 + \cos \theta) \quad (2)$$

The apparent separation velocity is the sum of these

$$v_s = 2c \sin \theta/(1 - \cos^2 \theta) = 2c/\sin \theta \quad (3)$$

Note that the apparent velocity of separation can not be  $< 2c$ . Also if  $S_1$ ,  $S_2$  and O are all seen, then the separation ratio

$$\frac{S_1 O}{S_2 O} = \frac{1 + \cos \theta}{1 - \cos \theta} \quad (4)$$

and so  $\theta$  can be determined direct from such observations. With  $\theta$  known and the apparent angular separation rate  $\dot{\chi} = v_s/D$  found from VLB observation then the distance  $D$  is  $v_s/\dot{\chi} = 2c(\dot{\chi} \sin \theta)^{-1}$ . The Hubble constant is then determined from the redshift  $z$  of the object at O by  $H_0 = cz/D$ .

### Modification factors

The above theory is subject to simple modification when the object O is not observed at rest but nearby at a large redshift  $z$ . In the first case the observer allowing for time dilation attributes an 'apparent' expansion velocity  $v_s$  related to  $\dot{\chi}$  by  $\dot{\chi} = v_s/[D(1+z)]$ . The events seem to happen slower than they really do because of the redshift. Second, the great time lapse between the emission and the reception of the radiation is large and the geometry may not be Euclidean or Minkowskian. In particular in Friedmann universes the formula connecting the apparent transverse motion  $\dot{\chi}$  and  $v_s$  is not

$$v_s = D\dot{\chi}(1+z) \div \frac{cz}{H_0}(1+z)\dot{\chi} \quad (5)$$

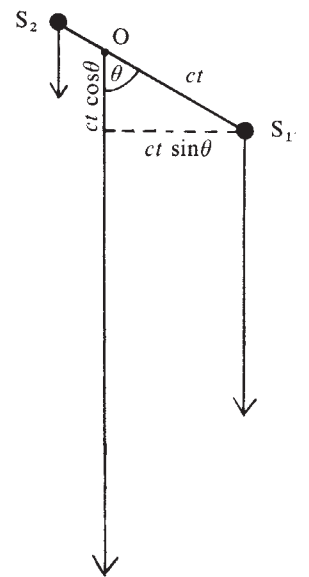
where  $H_0$  is the Hubble constant, but is modified by factors dependent on the acceleration parameter  $q_0$ . The full formula (given in ref. 5) is

$$\frac{v_s}{c} = \frac{\dot{\chi}}{H_0} q_0^{-2} (1+z)^{-1} \{q_0 z - (1-q_0)[(1+2q_0 z)^{1/2} - 1]\} \quad (6)$$

for  $q_0 z \ll \frac{1}{2}$  this gives

$$\frac{v_s}{c} = \frac{z}{H_0} (1+z)^{-1} [1 + \frac{1}{2}(1-q_0)z] \quad (7)$$

Note the correction factor of  $(1+z)^{-2}[1 + \frac{1}{2}(1-q_0)z]$  as compared with the simple equation (5). Equation (4) being a ratio of separations still holds good as do equations for  $v_1/v_s$  and  $v_2/v_s$  obtained by dividing equation (1) by (3) and equation (2) by (3).



**Fig. 1** A flash at O in a tube illuminates and excites two patches of plasma seen by the antenna A at positions  $S_1$  and  $S_2$ . Whereas both  $S_1$  and  $S_2$  actually travel out at the speed of light,  $S_2$  is seen at an earlier source epoch because it is further away.



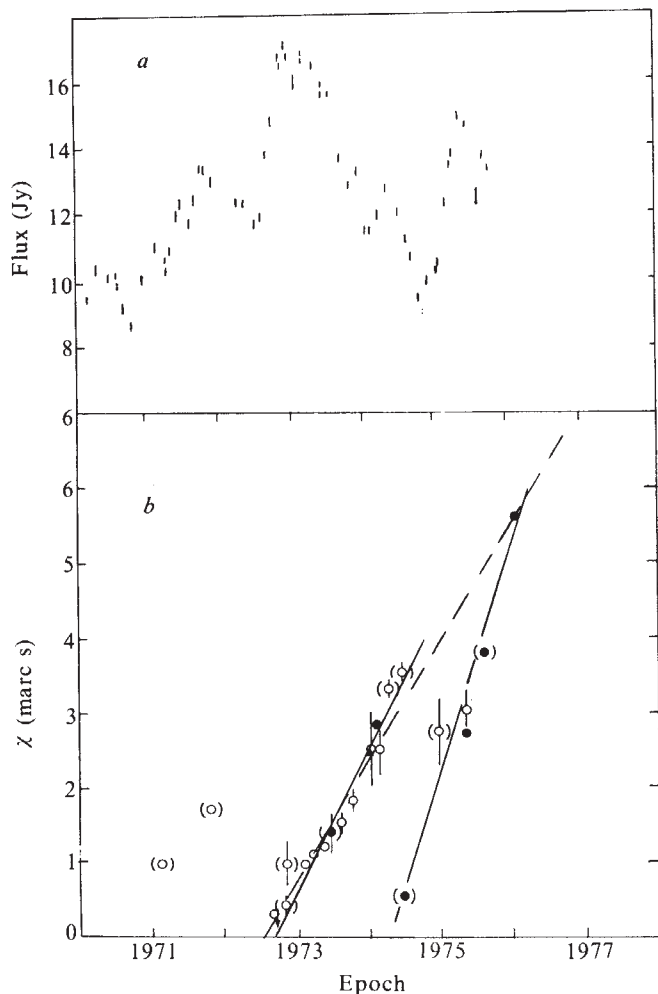


Fig. 2 The observations of 3C120 as reported in ref. 5 but with a dotted line superposed. *b*, plots  $\chi$  the separation angle  $S_1AS_2$  against time. ●, 2.8 cm; ○, 3.8 cm; ▲, 6.0 cm. *a*, The 3.8 cm flux.

### Application to 3C120

Figure 2 shows the observations given in ref. (5). The 3.8-cm flux shows an outburst in 1973 which caused a double source. This was seen as a simple double expanding at about  $v_s = 5c$  (solid line) until 1974.2 when the double model failed to fit well, because of a small increase of flux. In 1975 there followed a violent increase of flux and immediately a two source model fitted well again, but at about the same separation as it was at 1974.0. The source began to dim in late 1975 and in 1976.1 it was again a well-determined double although the flux then was not given. Cohen *et al.* interpreted these events as the two super-luminal expansions of  $5c$  and  $8c$  given by the heavy straight lines. In my interpretation, however, this would involve the vortex tube up the rotation axis of the radio galaxy to turn between the two events and I consider this unlikely, especially as the observed position angle of separation on the sky are the same for both doubles. The following interpretation of events is more natural. 1972.5–1973 a double expanding source  $S_1, S_2$  was born at O and expanded throughout 1973 as its flux diminished. In early 1974, the object at O brightens, confusing the neat double interpretation of the data. In 1975.2, O has increased in brightness so much that it forms a double with  $S_1$ , the source moving towards the Earth and  $S_2$  is outshone and ignored. O dims in late 1975 leading to a confused picture with  $S_1$ , O and  $S_2$  all seen. Finally in 1976, O dies away leaving the old expanding source  $S_1, S_2$ . This latest point fits on the dotted expansion line ( $4.6c$ ) of the old source, which is a good alternative to the full line ( $5c$ ) if one ignores confused data. The importance of this new interpretation arises because it not only removes the discrepancy of two different super-luminal velocities observed from the same object, and therefore probably at the same  $\theta$  but, because we see O at times we can

measure the ratio of the separations  $OS_1$  to  $OS_2$  and hence determine  $\theta$ . Hence, the apparent transverse separation velocity is known from equation (3) so the observed angular velocity gives the distance and hence Hubble's constant. The small  $z = 0.033$  of 3C120 makes the cosmological corrections unimportant for this source, so the Hubble constant derived in this case is independent of  $q_0$ . The same method applied to 3C345 should also yield a good value of  $q_0$ .

If  $\theta$  were  $90^\circ$  the clear double seen in 1975.4 would have been half way up to the dotted line. Its fractional offset  $[OS_1 - \frac{1}{2}S_1S_2]/S_1S_2$  corresponds to the offset of O from the centre of  $S_1$  and  $S_2$  and is, therefore,  $\frac{1}{2} \cos \theta$ , from this  $\theta = 68^\circ \pm 10^\circ$ . It is particularly interesting that this angle is close to the inclination of the apparent disk of 3C120 measured from Arp's picture on which it seems to be a disturbed spiral<sup>12</sup>. From his picture,  $i = 55\text{--}60^\circ$  but this may be in error by  $10^\circ$  due to the fact that the disturbed disk does not have a circular outline. Furthermore the observed separation of the radio double, although fairly close to the apparent minor axis of the optical image is not along it. It seems reasonable to assume that ejection is taking place along the rotation axis, but the galaxy has imperfect circular symmetry. In such a case  $i$  should equal  $\theta$  and as both determinations have similar errors we shall take  $\theta = 64 \pm 8^\circ$ . The light echo theory then gives

$$v_s = 2c/\sin 64^\circ = 2.22c \pm 0.19c.$$

Cohen *et al.* adopt  $H_0 = 55 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and obtained  $5c$  for this velocity, but our correction to the dotted line interpretation of Fig. 2 yields  $4.6c$  on that scale. To deduce the answer from the light echo theory the Hubble constant must be changed by  $(4.6 \pm 0.2)/(2.22 \pm 0.19) = 2.1 \pm 0.2$  producing Hubble constant of  $115 \pm 12 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

### Application to 3C273

Triple models have been given<sup>13</sup> for 3C273 and the best ratio of source distances from the centre gives  $\cos \theta = 0.2 \pm 0.2$ ,  $\theta = 78.5^\circ \pm 12^\circ$ . The expected expansion is  $2c/\sin \theta = 2.04c$  while the observations give  $4.2c$  on the  $H_0 = 55$  scale, with  $q_0 = 0.05$ . The correction factor is thus  $4.2/2.04 = 2.05$  yielding a Hubble constant of  $113 \pm 10 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

### Statistics of super-luminal expansions

From equation (3)  $v_s = 2c/\sin \theta$  so this velocity cannot be less than  $2c$ . Furthermore most of the solid angle of a sphere is not too far from its equator, so  $v_s$  ought often to be just greater than  $2c$ , whereas  $v_s > 6c$  should occur only seldom. The times a super-luminal velocity less than some chosen  $v_s$  is seen should be

$$\frac{N(v_s)}{N_0} = (2\pi)^{-1}[(2\pi - 2\pi(1 - \cos \theta))] = \cos \theta = \left[1 - \left(\frac{2c}{v_s}\right)^2\right]^{1/2} \quad (8)$$

The distribution function for  $v_s$  should be  $-d(N/N_0)/dv_s$  and this is plotted as Fig. 3. With only three well determined super-luminal velocities to consider, however, it is better to consider the formulae for the integrated distribution and its moments. In particular averaged over angle

$$\left\langle \frac{1}{v_s} \right\rangle = \int \frac{1}{v_s} \frac{dN}{N_0} = \frac{\pi}{8c} = \frac{0.79}{2c} \quad (9)$$

and

$$\sigma^2 = \left\langle \left( \frac{1}{v_s} - \left\langle \frac{1}{v_s} \right\rangle \right)^2 \right\rangle = \left( \frac{2}{3} - \frac{\pi^2}{16} \right) \frac{1}{4c^2} \quad (10)$$

$$\sigma = 0.22/2c.$$

The three well determined super-luminal velocities are, on the  $H_0 = 55 \text{ km s}^{-1} \text{ Mpc}^{-1}$  scale with  $q_0$  assumed small ( $q_0 = 0.05$ ),  $v_s/c = 4.6, 4.2$ , and  $7.0$  in the objects 3C120, 3C273 and 3C345 respectively. Thus with standard errors for the mean

$$\left\langle \frac{1}{v_s} \right\rangle = (0.20 \pm 0.05) \frac{1}{c} = (0.40 \pm 0.10) \frac{1}{2c}$$

To scale this to the theoretical value of equation (9) we must change

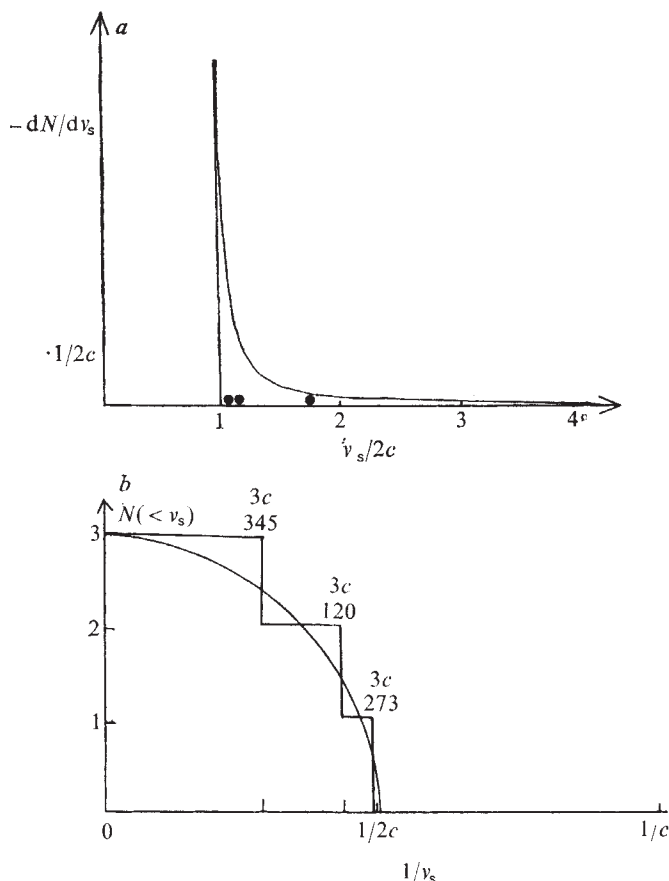


Fig. 3 a, The distribution of superluminal velocities expected from the light echo theory, and b, its integral plotted against  $2c/v_s$  compared with the histogram of observed data.

Hubble's constant by  $0.79/(0.4 \pm 0.1) = 2.0 \pm 0.5$  giving  $H_0 = 110 \pm 30$ . One further constraint on the factor may be independently deduced from the expansion of 3C273 at  $4.2c$  on the  $H_0 = 55$  scale. From our theory this must be greater than  $2c$  so the correction factor cannot be larger than 2.1. Taking this and its error into account the best value of the correction factor is  $2.0 \pm 0.1$  and the Hubble constant is  $110 \text{ km s}^{-1} \text{ Mpc}^{-1}$  with a maximum error upwards of  $10 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and an r.m.s. error downwards of about 10 also.

Principles for interpreting super-luminal source data on light

echo theory are (1) Bursts occur at the source O. (2) They usually cause pairs of expanding sources  $S_1, S_2$  of which  $S_1$ , the one nearer the observer, is normally seen as the brighter. Since  $S_2$  is seen at an earlier source epoch it may have a flatter (harder) spectrum than  $S_1$ . (3)  $S_1$  and  $S_2$  expand outwards at velocities  $c \sin \theta / [1 - \cos \theta]$  and  $c \sin \theta / [1 + \cos \theta]$  respectively. These velocities do not accelerate. (4) Further bursts may occur at O leading to a brightening of O possibly followed by further source pairs  $S_1' S_2'$  travelling along the same position angle and at the same velocities as  $S_1$  and  $S_2$  respectively. (5) All source pairs will eventually fade but the source O is responsible for all very rapid time variability.

## Predictions

The light echo interpretation of the 3C120 data suggest that when the 3.8 cm flux at 1976.1 is disclosed it will again be low. Also that the next experiment will show a point on the dotted line of Fig. 2 if the flux continues low. It also suggests that the ratio  $S_2 O / S_1 O$  will be about 0.10 if O is seen in 3C345 (we may already have seen O and not  $S_2$  in which case we still expect the same ratio).

The new value of  $H_0$  predicts that no super-light velocities  $v_s$  less than  $4c$  on the old  $H_0 = 55$  scale will be detected. If, however,  $S_2$  were not seen the separation speeds of  $v_1$  down to  $2c$  on that old scale could be found. Finally each source should give the same value for  $H_0$  if this interpretation is correct.

## Conclusions

The Hubble constant is  $110 \pm 10 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . The universe is half the linear size and half the age that would be attributed to it on the old Hubble constant of  $55 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . The new  $H_0^{-1}$  age is  $9 \times 10^9$  yr and this means that Einstein-de Sitter age of  $\frac{2}{3}H_0^{-1}$  is embarrassingly small for stellar evolution theory. Hence, it is probable that the Universe is open with a small value of  $q_0$ . For many years Van den Bergh and de Vaucouleurs, and more recently Dibai, Hanes, Hartwick and Madore, have advocated almost as large values of the Hubble constant.

I thank the large VLBI teams for their hard work and J. R. Shakeshaft, D. Richstone and G. Wegner for helpful discussions.

Received 3 August; accepted 3 October 1977.

1. Rees, M. J. *Nature* **211**, 468 (1966).
2. Blandford, R. D., McKee, C. F. & Rees, M. J. *Nature* **267**, 211 (1977).
3. Perrine, C. D. *Astrophys. J.* **17**, 310 (1903).
4. Russell, H. N., Dugan, R. S. & Stewart, J. Q., *Astronomy* **2**, 786 (Ginn & Co., Boston, 1927).
5. Cohen, M. H. et al. *Nature* **259**, 17-20 (1976).
6. Lynden-Bell, D. *Proc. Copenhagen Quasar Conf.* (1977).
7. Lynden-Bell, D. *Nature* **223**, 690 (1969).
8. Lynden-Bell, D. *Vatican Symp. Nuclei of Galaxies* 527 (ed. O'Connell, D.) (Pont. Acad. Sci., 1970).
9. Lynden-Bell, D. & Rees, M. J. *Mon. Not. R. astr. Soc.* **175**, 613 (1971).
10. Bardeen, J. M. *Nature* **226**, 64 (1970).
11. Shakura, N. I. & Sunyaev, R. A. *Mon. Not. R. astr. Soc.* **175**, 613 (1976).
12. Arp, H. *Pub. astr. Soc. Pac.* **87**, 546 (1976).
13. Schilizzi, R. T. et al. *Astrophys. J.* **201**, 263 (1975).

# Optimum efficiency of photogalvanic cells for solar energy conversion

W. John Albery

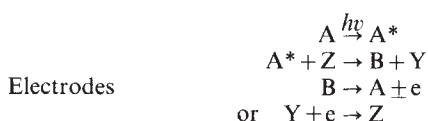
Physical Chemistry Laboratory, South Parks Road, Oxford, UK

The performance of photogalvanic cells for the direct conversion of solar energy to electrical energy depends on the cell photochemistry, the homogeneous kinetics, the mass transport, the electrode kinetics and the load on the cell. The variation of the power output with the concentrations of the redox couples, their transport and kinetic parameters and the dimensions of the cell is found. The power conversion efficiency of the optimal cell could be as large as 18% but it is unlikely that all the necessary conditions can be met. A more realistic estimate of the maximum power conversion efficiency that could be achieved from a photogalvanic cell is between 5 and 9%.

Mary D. Archer

Department of Physical Chemistry, Lensfield Road, Cambridge, UK

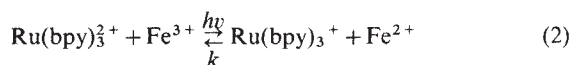
PHOTOGALVANIC cells are possible devices for the direct conversion of solar energy into electrical energy. This paper is concerned with the maximum power that can be achieved by such devices and with describing the conditions for the most efficient operation of such cells. The cell works by the reactions



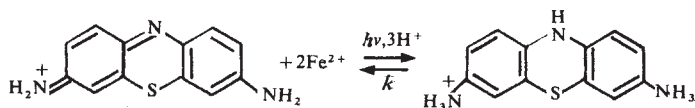
(1)



Two redox couples A,B and Y,Z are present in the solution; in the dark there is a very small concentration of B. On illumination, the absorption of radiation by A and the subsequent electron transfer reaction of A\* with Z generates B. The photogenerated B reacts on one of the electrodes thereby generating power. Examples of this type of system are shown in equations (2) and (3). We assume that the concentrations of Y and Z are large, compared to the concentration of A so that the concentrations of Y and Z throughout the cell are not significantly perturbed from their values in the dark.



where bpy = 2,2'-bipyridine.



The best geometry for the cell consists of two parallel electrodes separated by a thin layer (~0.1 mm) of electrolyte solution. One of the electrodes is transparent (such as  $\text{SnO}_2$ ); the light is shone through this electrode, as shown in Fig. 1.

### Reactions on electrodes

The performance of the cell depends first on the kinetics of the electrode reactions. On either electrode, either couple can have 'reversible' or 'irreversible' electrode kinetics. When a couple is reversible the electrode kinetics are rapid and the surface concentrations of the couple at the electrode obey the Nernst equation. When a couple is irreversible the electrode kinetics are sluggish and we only consider the extreme case where no electrode reaction takes place. The performance of cells, in which couples have intermediate electrode kinetics, will be between the two extremes that we consider. Four of the 16 possible cases are shown in Table 1.

Eight cases which have irreversible electrode kinetics for both couples on either of the electrodes can be rejected because to pass current it is essential that one of the couples should be reversible on each electrode. Four cases in which the Y,Z couple is reversible on both electrodes can be rejected because we have assumed that there is an insignificant perturbation to the concentrations of the Y,Z couple; so if the couple is reversible, the two electrodes will have the same potential and no power is produced. Similarly the fourth case of Table 1 can be rejected because at the dark electrode the reversible A,B couple will give the same potential as the reversible Y,Z couple at the illuminated electrode. This leaves three cases worth considering. In the first, the electrode kinetics are similar on the two electrodes and hence they can be made of the same material. The cell works as a 'concentration cell'. In the

other two cases the electrode kinetics are different on the two electrodes and the cell works with 'differential electrode kinetics'.

### Concentration of B in solution

To find the power developed by the cell we have to calculate how the concentration of B varies with the distance between the two electrodes. A steady state is established described by the differential equation (4)

$$D \frac{\partial^2 b}{\partial x^2} + \phi I \epsilon a - kby = 0 \quad (4)$$

where  $D$  is the diffusion coefficient of B;  $x$  describes the distance from the illuminated electrode;  $\phi$  is the quantum efficiency for the generation of B from A;  $I$  is the quantum intensity of the light;  $\epsilon$  is the natural extinction coefficient of A;  $a$ ,  $b$ , and  $y$  are the concentrations of A, B and Y respectively; and  $a_D$  is the concentration of A in the unilluminated solution.

The first term describes the diffusion of B, the second term its photochemical generation and the third term the thermal reaction with Y. There is no term describing transport by convection because we assume that the electrodes are close enough (< 0.1 mm) for natural convection not to develop. There is no advantage in having the electrodes further apart and there may be the disadvantage that the internal resistance of the cell may be larger.

The light intensity in equation (4) varies with  $x$  and obeys

$$\frac{\partial I}{\partial x} = -\epsilon a I = -\epsilon(a_D - b)I \quad (5)$$

where we have assumed that the diffusion coefficients of A and B are equal so that  $a + b = a_D$ . This differential equation gives the Beer-Lambert law when  $b \ll a_D$ , but we have also investigated the situation where  $b \approx a_D$  which means that the solution is bleached.

The boundary condition for equation (5) is that at  $x = 0$ ,  $I = I_0$ .

The boundary conditions for equation (4) depend on the electrode kinetics. At the illuminated electrode for the first three cases in Table 1

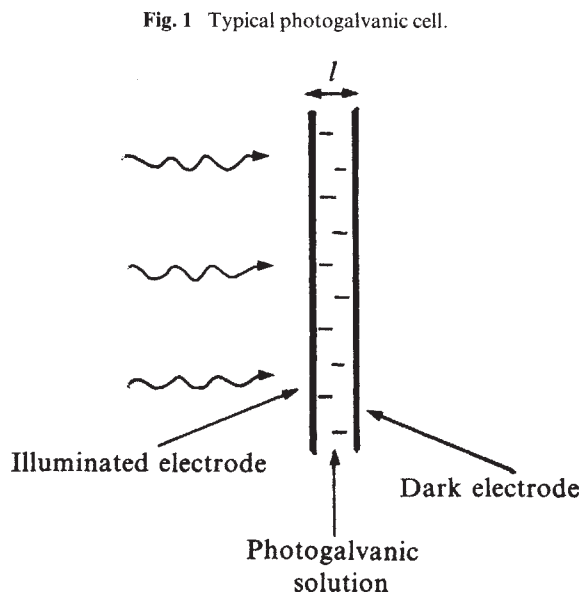
$$\left( \frac{\partial b}{\partial x} \right)_0 = b'_0 = \frac{|i|}{AFD} \quad (6)$$

At the dark electrode the boundary conditions depend on the different electrode kinetics shown in Table 1 and are: concentration cell,  $(\delta b / \delta x)_1 = b'_1$ ; differential electrode kinetics case I  $(\partial b / \partial x)_1 = 0$ ; case II  $b_1 \rightarrow 0$ .

The case II boundary condition arises because with both couples reversible, the dark electrode re-establishes at its surface

Table 1 Reversible (R) or irreversible (I) electrode kinetics

| Couple | Illuminated electrode |     | Dark electrode |     | Notes                              |
|--------|-----------------------|-----|----------------|-----|------------------------------------|
|        | A,B                   | Y,Z | A,B            | Y,Z |                                    |
| Case 1 | R                     | I   | R              | I   | Concentration cell                 |
| Case 2 | R                     | I   | I              | R   | Differential electrode kinetics I  |
| Case 3 | R                     | I   | R              | R   | Differential electrode kinetics II |
| Case 4 | R                     | R   | R              | I   | Reject                             |



the equilibrium values of the concentrations found in the unilluminated solution; any B is destroyed by the electrode:



The solution of the differential equations with the boundary conditions has been described elsewhere<sup>1,2</sup>. The value of  $b$  at the illuminated electrode allows the potential at that electrode to be calculated from the Nernst equation, as in all three cases the A,B couple is reversible. For the concentration cell a similar calculation is carried out at the dark electrode. For cells with differential electrode kinetics the potential difference at the dark electrode is unchanged from its value where the cell is unilluminated. Hence for each value of the current using equation (6) we can calculate the potential difference across the cell and hence the power produced.

## Results

In giving the results of the calculations it is convenient to introduce the characteristic lengths given in Table 2. In Fig. 2 we show some typical concentration profiles for different values of the characteristic lengths and for the different electrode kinetics given in Table 1. The profiles are drawn for the case where the cell is short circuited. We also show how the light is absorbed in the cell. In all these cases the solution is not bleached and the generating length is unimportant. The current at short circuit is a maximum if the following condition holds:

$$X_e < X_k + X_l \quad (8)$$

The photoelectrochemical collection efficiency,  $N_{hv}$  (ref. 3) describes the flux of electrons from the cell divided by the incident flux of photons. For the conditions given by equation (8) it is equal to its maximum value of unity for a cell with differential electrode kinetics, (cases 2 and 3 of Table 1). All the photons are trapped close to the illuminated electrode and turned into current. For the concentration cell the maximum value of  $N_{hv}$  is  $\frac{1}{2}$  (S. W. Feldberg, personal communication). Even though the photons are trapped close to the illuminated electrode, the fact that the Y,Z couple is not electroactive on the dark electrode means that at short circuit the concentration of B has to be equal at the two electrodes. This in turn means that the photogenerated flux of B partitions equally between diffusion away from the illuminated electrode and diffusion away from the dark electrode after reaction at both electrodes.

Although the condition given in equation (8) maximises the current developed by the cell, the power also depends on the voltage at which the current is produced. The maximum power is found for cells in the hatched region of Fig. 2. For concentration cells it is essential to have the cell length,  $X_l$ , greater than the reaction length,  $X_k$ , or else B can diffuse all over the cell producing a uniform concentration profile and hence very little power. For the cells with differential electrode kinetics, increasing the concentration of Y increases the potential difference developed by the cell through the effect of Y on the potential of the Y,Z half cell at the dark electrode. But increasing the concentration of Y also increases the rate of the back reaction  $B + Y$  and hence decreases

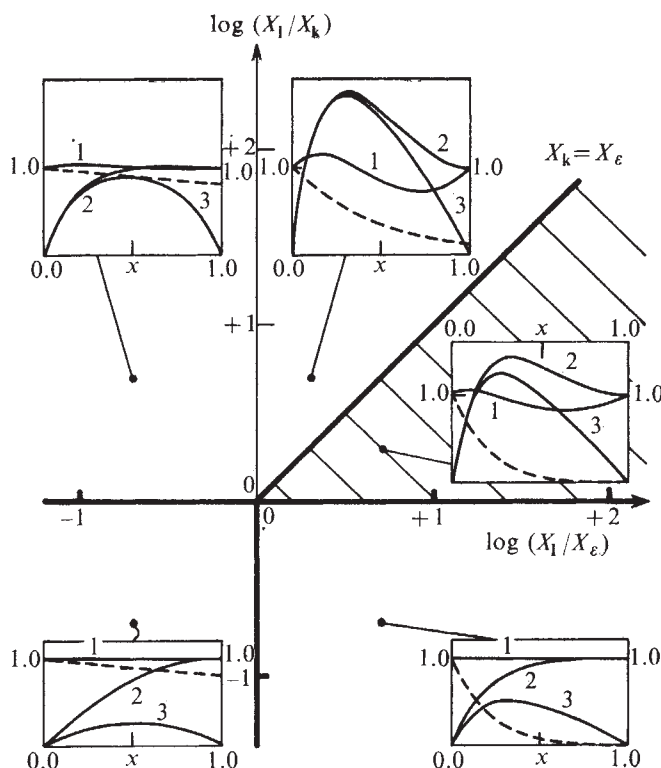


Fig. 2 Typical concentration profiles across photogalvanic cells. The illuminated electrode is on the left and the broken lines show the penetration of the light; the abscissa compares  $X_l$ , the cell length, with  $X_e$ , the absorbance length. The ordinate compares  $X_l$  with  $X_k$ , the reaction length, which describes the decomposition of the photo-generated intermediate B. The solid lines show the concentration of B for the first three cases in Table 1, when the cell is short circuited. In all three cases B reacts on the illuminated electrode. On the dark electrode B is generated, does not react and is destroyed for cases 1, 2 and 3 respectively. The cells are most efficient in the hatched region.

$X_k$ . Thus the optimum power point for this type of cell depends on the concentration of Y and is found in the hatched area close to the line  $X_k = X_e$ . Detailed investigation of the cases where the dye, A, is bleached shows that they produce less power than cases in the hatched region of Fig. 2. The voltage developed by such cells is larger but the current is very much less because the photons are absorbed too far away from the illuminated electrode for the photogenerated B to reach it before decomposing.

The maximum power  $W_{max}$  produced by the concentration cell<sup>1</sup> can be shown to be

$$W_{max} = 0.28 ART\phi I_0 \quad (9)$$

where  $A$  is the area of each electrode.

This is the maximum possible power that can be obtained after optimising the concentrations of the species, the dimensions of the cell and the load on the cell. The conditions for this maximum are:

$$X_e < X_G < X_k < X_l \quad (10)$$

and

$$X_G^3 = X_k^2 X_e \quad (11)$$

The maximum power produced by a cell with differential electrode kinetics can be shown to be<sup>2</sup>:

$$W_{max} = 0.8 AF\phi I_0 [|\Delta E| + \frac{RT}{F} \{ \ln \frac{\phi I_0 e}{k[Z]} - 1.6 \}] \quad (12)$$

$$\simeq 0.8 AF\phi I_0 \Delta E \quad (13)$$

where  $|\Delta E|$  is the difference in standard electrode potentials for the A,B and Y,Z couples.

Table 2 Characteristic lengths

| Name              | Symbol and equation          | Description                                                                           |
|-------------------|------------------------------|---------------------------------------------------------------------------------------|
| Cell length       | $X_l$                        | Distance between electrodes                                                           |
| Absorbance length | $X_e = (e\alpha_0)^{-1}$     | Unless solution is bleached, the light is absorbed mainly in this distance            |
| Generating length | $X_G = (D/\phi e I_0)^{1/2}$ | Average distance A diffuses in light of intensity $I_0$ before being converted into B |
| Reaction length   | $X_k = (D/k_y)^{1/2}$        | Average distance B diffuses before being converted to A                               |



The conditions for this maximum are:

$$10 X_e \simeq X_G \simeq \frac{1}{2} X_k < X_l \quad (14)$$

If these conditions are fulfilled then equation (13) describes the maximum power from cells with differential electrode kinetics whether they be case 2 or 3 of Table 1.

## Discussion

For both types of cell we find the same order for the characteristic lengths in equations (10) and (14). The shortest length must be  $X_e$  to absorb all the photons close to the electrode. The inequality  $X_k > X_e$  ensures that the B, that is formed within the distance  $X_e$ , is not decomposed by reaction with Y before it reaches the electrode. The conditions for  $X_G$  ensure that the solution does not become bleached near the illuminated electrode. The condition for  $X_l$ , the distance between the electrodes, means that the dark electrode does not interfere with the trapping of the photons and the reaction of B on the illuminated electrode.

The maximum power that the concentration cell can deliver is given by equation (9). For the maximum power point the current is close to its short circuit value; the collection efficiency is  $0.3\phi$ , and the current density is  $0.3\phi I_0$ . But the cell voltage is only about  $(RT/F)$  volts. Since this is a concentration cell, the potential difference arises from the  $RT/F \ln$  term in the Nernst equation.

One cannot draw significant currents from a concentration cell and at the same time maintain very different concentrations at the two electrodes.

To calculate power outputs from equations (9) and (13) we must select a reasonable value for  $I_0$ . In AM2 solar radiation (of total irradiance  $749 \text{ W m}^{-2}$ ),  $I_0$  values for photons of wavelengths less than 700, 600 and 500 nm are  $1.57 \times 10^{-3}$ ,  $9.46 \times 10^{-4}$  and  $4.2 \times 10^{-4} \text{ mol photons m}^{-2} \text{ s}^{-1}$  respectively<sup>4</sup>. We shall take a value of  $1.6 \times 10^{-3} \text{ mol photons m}^{-2} \text{ s}^{-1}$  as a reasonable estimate of the flux density of photochemically active photons in AM2 sunlight. As relative values of solar spectral irradiance do not vary very greatly with atmospheric conditions<sup>5</sup>, conversion efficiencies calculated from these figures do not vary by more than a few per cent from efficiencies calculated from other solar spectral distributions; absolute outputs, of course, vary with the incoming total irradiance.

Taking  $I_0 = 1.6 \times 10^{-3} \text{ mol photons m}^{-2} \text{ s}^{-1}$  and  $\phi = 1$  we obtain from equation (9)

$$W_{\max}/A = 1.1 \text{ W m}^{-2} \quad (15)$$

This is the maximum power that can be obtained from a photogalvanic concentration cell driven by solar energy. The power conversion efficiency is no more than 0.15% and we conclude that the concentration cell cannot be an effective device for solar energy conversion.

For a cell with differential electrode kinetics, the maximum power is given by equation (13). Not only is the collection efficiency much closer to unity but also the cell voltage is now given by  $\Delta E$  rather than the miserable  $RT/F$  of the concentration cell. The optimum value for  $\Delta E$  is 1.1 V<sup>6</sup>. With this value and the same value of  $I_0$  as above we obtain

$$W_{\max}/A = 140 \text{ W m}^{-2} \quad (16)$$

This corresponds to a power conversion efficiency of 18%. In principle the optimum case for this type of cell is almost as efficient as a silicon solar cell. Nearly all the photons that can drive the reaction are trapped. While in the ideal case negligible energy would be wasted between the absorption of a photon with threshold energy and the formation of B, in the case we have taken the threshold photon energy is 700 nm or 1.8 eV and the flux of electrons is produced at the lesser potential difference of 1.1 V. Had we taken threshold photon energies of 600 or 500 nm the conversion efficiencies would have been 11% and 5% respectively. To achieve this amount of power we have to satisfy equation (14),

however. For typical values of  $D = 10^{-5} \text{ cm}^2 \text{ s}^{-1}$  and  $\varepsilon = 10^8 \text{ cm}^2 \text{ mol}^{-1}$  we find

$$X_G \sim 10^{-3} \text{ cm} \quad (17)$$

This means that

$$X_e \sim 10^{-4} \text{ cm} \quad (18)$$

or

$$[A] \sim 10^{-1} \text{ M} \quad (19)$$

Hence the dye, A, has to be rather soluble. Furthermore, the electrochemical rate constant  $k'$ , for the destruction of B on the illuminated electrode must satisfy the condition

$$k' > D/X_e \sim 10^{-1} \text{ cm s}^{-1} \quad (20)$$

If this condition is not satisfied the electrode will not remove all the B that reaches it and some B will be lost by reaction with Y in the bulk of the solution. This is a severe condition as most electrochemical rate constants are less than  $10^{-2} \text{ cm s}^{-1}$ . The condition in equation (14) with the value for  $X_G$  in equation (17) means that the kinetics of the reaction of B with Y must satisfy the condition,

$$k[Y] \sim 40 \text{ s}^{-1} \quad (21)$$

But the concentration of Y must be at least twice the concentration of B generated in the light. This is to ensure that there is no concentration polarisation at the dark electrode. Hence for the optimum case

$$[Y] > 2[B] \sim 10^{-2}[A] \sim 10^{-3} \text{ M} \quad (22)$$

Substitution in equation (21) gives  $k < 4 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ . Thus from equations (19) to (21) we may summarise the contradictory requirements for the optimum case: (1) the species, A, must be rather soluble,  $[A] > 0.1 \text{ M}$ ; (2) the electron transfer reaction of the A,B couple on the illuminated electrode must be very rapid,  $k' > 10^{-1} \text{ cm s}^{-1}$ ; (3) considering the thermodynamic driving force corresponding to  $\Delta E$  of 1.1 V, the electron transfer reaction of B + Y must be sluggish:  $k < 4 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ ; (4) the electron transfer reaction that forms B and Y from  $A^* + Z$  must be rapid. Electron transfer reactions that obey the Marcus theory<sup>7</sup> are unlikely to satisfy all the conditions 2, 3 and 4 (S. W. Feldberg, personal communication).

The detailed treatment presented elsewhere<sup>2</sup> allows the optimum conditions and the power that can be produced for any particular system to be calculated. For instance, if we relax the condition of equation (14), so that  $X_e = X_G = X_k$  then we find that  $[A] \sim 10^{-2} \text{ M}$ ;  $k' > 10^{-2} \text{ cm s}^{-1}$ ;  $k < 4 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$ .

It may be possible to find a system that satisfies these conditions. For this case the power is only reduced by a factor of 2 from the optimum to about 9% and

$$W_{\max}/A = 70 \text{ W m}^{-2} \quad (23)$$

It is clear that the processes occurring in a photogalvanic cell are controlled by the photochemistry, homogeneous kinetics, mass transfer and electrochemical kinetics of the system. The power developed depends on the concentrations of all the species, the intensity of illumination, diffusion lengths at the electrode, the electrode kinetics, the spacing of the electrodes and the load on the cell. The optimum performance of the cell can only be obtained by carefully selecting the best values for all these variables.

We thank Dr Stephen Feldberg for helpful discussions, and Mr Stephen Fisher and Mr Andrew Foulds for their criticisms.

Received 3 August; accepted 1 October 1977.

1. Albery, W. J. & Archer, M. D. *J. electroanal. Chem.* (in the press).
2. Albery, W. J. & Archer, M. D. *J. electroanal. Chem.* (in the press).
3. Albery, W. J., Archer, M. D. & Egdell, R. G. *J. electroanal. Chem.* **82**, 199 (1977).
4. Data from proposed AM2 spectral distribution in *Interim Solar Cell Testing Procedures for Terrestrial Applications* 13 (NASA, 1975).
5. Gates, D. M. *Science* **151**, 523 (1966).
6. Archer, M. D. *Solar Energy* (in the press).
7. Marcus, R. A. *J. chem. Phys.* **43**, 679 (1965).

# Cordilleran Benioff zones

Peter J. Coney & Stephen J. Reynolds

Department of Geosciences, University of Arizona, Tucson, Arizona 85721

*Ages of late Mesozoic-early Cainozoic igneous rocks in south-western North America indicate a magmatic arc initiated near the continental margin, swept over 1,000 km northeastward, then swept back, all in 110 Myr. The sweep in and return are interpreted as due to flattening of a Benioff zone to  $< 15^\circ$  followed by rapid collapse.*

THE geometric relationship between trench, dipping Benioff zone, and magmatic arc is well established in plate tectonics<sup>1-3</sup>. In active arc-trench systems, distance from the trench axis to the main volcanic front or magmatic arc axis varies from as little as 75 km to over 400 km<sup>4,5</sup>. Much of this variation is due to the width of the accretionary prism which seems to control the initial angle of descent of subducting slabs<sup>3,5</sup>. Once the slab is below a depth of 25-75 km the angle of descent usually steepens, but the descent angle of deeper segments of Benioff zones can range from  $90^\circ$  to as little as  $15^\circ$  (ref. 7). Surface manifestation of arc activity can extend over some width and stands between 100 km and 300 km above the Benioff zone<sup>6</sup>. Based on this geometry, the distance the magmatic arc activity can lie inboard from the trench is controlled by the dip-angle of the descending slab.

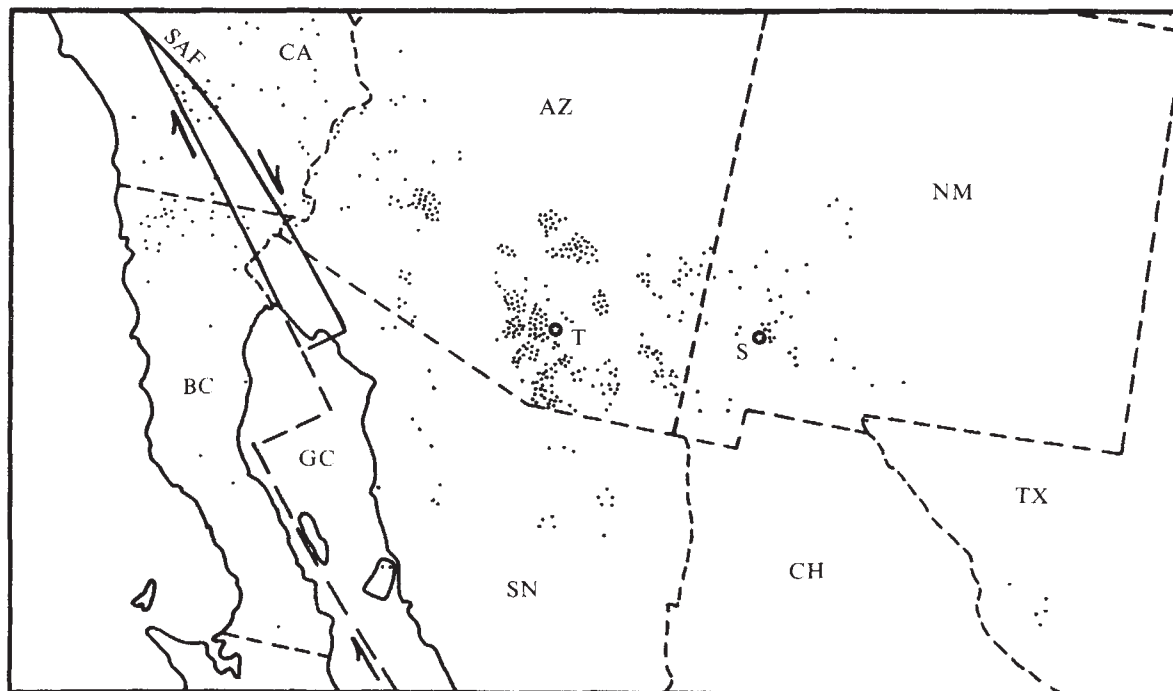
Extrapolation of this model back through Mesozoic time in the North American Cordillera<sup>8-10</sup> has not been universally accepted. Here, igneous activity of arc type reached inboard in excess of 1,000 km from assumed marginal trenches and many have doubted the existence of resulting low-dipping Benioff

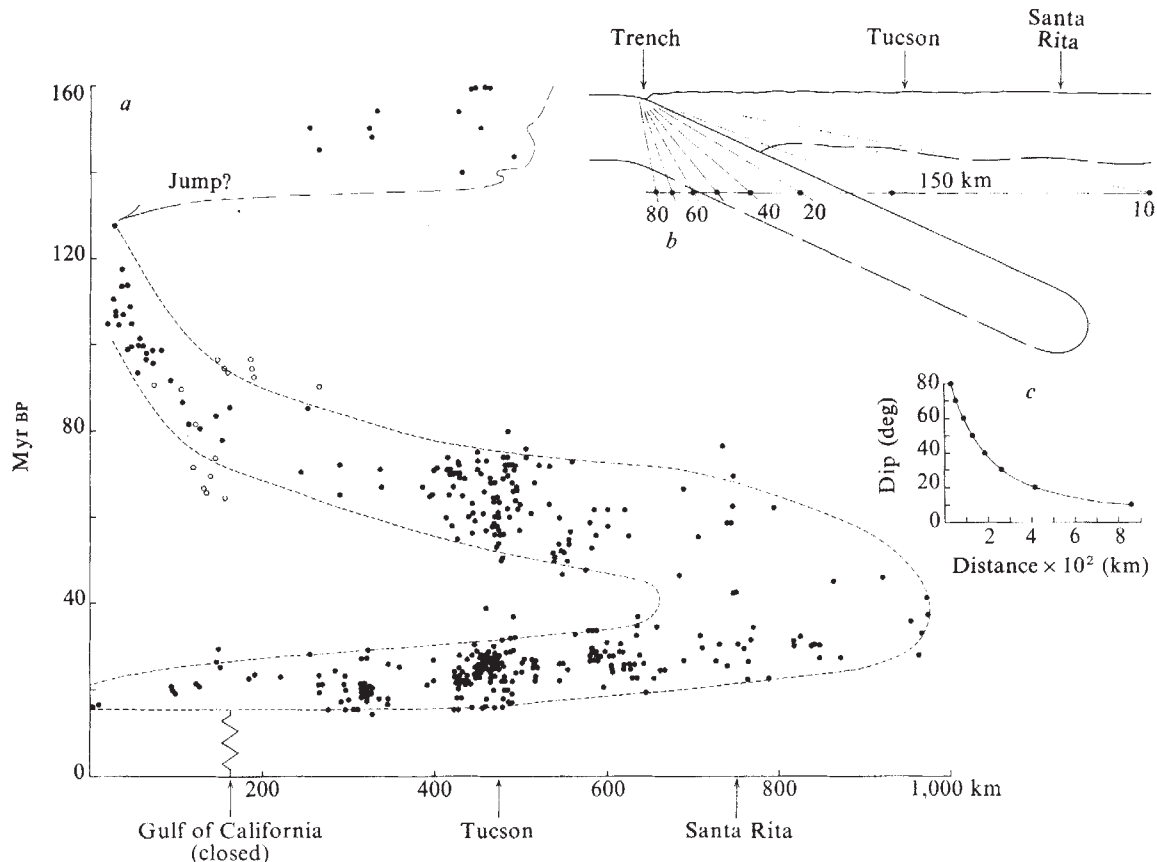
zone geometry<sup>11-13</sup>. We propose that widespread magmatic activity of arc type in southwestern North America during late Mesozoic to middle Tertiary time is consistent with Benioff zone geometry, and we show evidence that inferred changes in angle of Benioff zone dip can be tracked in time.

## Magmatic arc activity

Figure 1 shows the distribution of all available radiometric ages on igneous rocks from southern California, Arizona, New Mexico and west Texas in southwestern United States with a few additional ages from northern Baja California and Sonora in Mexico. The ages range from 160 Myr to about 15 Myr. The dated rocks have chemical compositions normally considered of magmatic arc type and include both volcanic and plutonic occurrences. The age dates are mostly by the K/Ar method supplemented with a few by Rb/Sr and U/Pb methods. Each dated occurrence has been examined for geological setting and method used. All available ages were used on Fig. 1 with the following exception: in the case of plutonic rocks from the Peninsular batholith in the western part of the region, multiple intrusion has caused extensive resetting of K/Ar ages<sup>14</sup>. As a result, only concordant biotite-hornblende pairs were used. In the absence of U/Pb ages concordant hornblende-biotite K/Ar pairs have been shown to be more reliable in belts of concentrated prolonged batholith emplacement<sup>15</sup>. Further east where plutonism was more dispersed single mineral dates were used. Throughout the region dates on volcanic rocks were considered accurate, providing they were geologically consistent and the method used indicated reliability.

**Fig. 1** Distribution of radiometric age determinations in southwestern United States and adjacent northwestern Mexico. Published dates from refs 14, 18-20, 44, 45. S. B. Keith and S. J. Reynolds compiled dates, lists and maps which are available from P. Damon, University of Arizona. Unpublished ages taken from refs 23, 44, and P. Damon, R. Nielson and W. Rehrig personal communication. T, Tucson; S, Santa Rita; SAF, San Andreas Fault; CA, California; AZ, Arizona; NM, New Mexico; BC, Baja California; GC, Gulf of California; SN, Sonora; CH, Chihuahua; TX, Texas.





**Fig. 2** *a*, Distribution of radiometric ages from Fig. 1 plotted as a function of time. Dates projected into a line passing through Tucson, Arizona, and Santa Rita, New Mexico after closure of Gulf of California. ○, single mineral K/Ar ages on plutons in southeastern California. *b*, Hypothetical arc-trench system scaled to southwestern North America. Various dip-angles of Benioff zones shown with points of intersection with 150 km depth line. *c*, Graphical representation of function controlling distance from trench to magmatic arc with varying Benioff zone dip angles.

Figure 2*a* shows the distribution of radiometric ages plotted as a function of time. To construct this figure, the Gulf of California was first closed, bringing dates south-west of the San Andreas Fault about 300 km southeastward relative to dates north-east of the fault. This reconstruction restores the age distribution to a pattern more like the original distribution before post-14 Myr opening of the Gulf of California and displacement on the San Andreas Fault. A line was then drawn from the western edge of the continent through Tucson, Arizona and Santa Rita, New Mexico, approximately perpendicular to the restored continental margin and the assumed late Mesozoic-mid-Tertiary trench. All the ages on Fig. 1 were then projected on to this line to produce Fig. 2*a*.

The scattering of ages at the top of Fig. 2*a* is the youngest part of a group of ages which represent a north-west-south-east trending magmatic arc across southern Arizona during Jurassic time<sup>18</sup>. This arc was extinguished in latest Jurassic time. Arc activity then seems to have jumped south-west towards the supposed site of the trench in early Cretaceous time. This jump established an arc near the continental margin and produced the Peninsular batholith of southern California-northern Baja California<sup>14,17</sup>.

Ages on the Peninsular batholith form a cluster at the far western edge of Fig. 2*a* ranging from 125 Myr to about 85 Myr. The dates shown are mostly K/Ar concordant biotite-hornblende pairs<sup>18-20</sup>, and they define an array which is oldest in the west and younger to the east across the batholith. Silver *et al.*<sup>17</sup> report a similar, but slightly older, trend in U/Pb ages. It can be seen from Fig. 2*a* that the magmatic pulse, apparently of just over 20 Myr duration, continues its sweep eastward, progressively decreasing in age. It eventually crosses the entire southwestern Cordillera, reaching central New Mexico 1,000

km inboard from the assumed trench about 40 to 55 Myr ago. Anderson and Silver<sup>21</sup> report a similar, but presumably slightly older, trend in U/Pb ages from Baja California across Sonora.

At its most easterly limit near 40 to 55 Myr, the magmatic pulse is scattered and somewhat diffuse. It then starts a rapid return migration, sweeping back across the same terrain over which it had earlier advanced, reaching southern California 15 to 20 Myr ago. Thus, in a period of just over 100 Myr, arc magmatism swept continuously inboard 1,000 km, then swept back to where it had originated. The dashed lines on Fig. 2*a* are a gross envelope superimposed on the data bracketing the magmatic arc distribution. Since most of the plutonic ages plotted are by the K/Ar method, they are believed to represent times when the igneous rocks passed through cooling temperatures somewhat less than 'emplacement' temperatures. We note, however, that the few published U/Pb ages fall within the envelope, and we assume that the average emplacement age for the magmatic arc at any given time probably falls somewhere within the envelope as well. At about 15 Myr, the character of igneous activity of the entire region analysed abruptly changed. The chemistry became distinctly bi-modal, with much basalt, and was associated with Basin and Range rifting<sup>22,23</sup>, initiation of movement on the San Andreas Fault, and eventual opening of the Gulf of California.

### Analysis and interpretation

Figure 2*b* shows a hypothetical arc-trench system scaled to a geometry likely to have existed in southwestern North America during mid-Cretaceous to mid-Tertiary time, with various Benioff zone dip-angles ranging from 80° to 10°. Also shown are distances inboard from the trench where the Benioff zone intersects the 150 km depth line. This depth is presumed to lie



below the locus of magmatic arc activity at the surface<sup>6</sup>. The function controlling the position of the magmatic arc axis with varying Benioff zone dip is: distance = depth  $\times$  cot  $\theta$ , where  $\theta$  is the dip. In other words, with decreasing dip the distance from the trench to the arc increases ever more rapidly Fig. 2c.

We suggest that the pattern of ages on Fig. 2a reflects the progressive flattening of a Benioff zone beneath southwestern North America from about 130–110 Myr to about 55–40 Myr, followed by a rapid steepening of dip from 40 Myr to about 15 Myr. This decreasing then increasing Benioff zone dip first swept a magmatic arc 1,000 km inboard from the trench until 55–40 Myr, then swept it back towards the continental margin by about 20 Myr.

Figure 3 is derived from Fig. 2a and shows the change in Benioff zone dip as a function of time beneath southwestern North America based on a best-fit curve through the centre of the envelope enclosing the age dates. Figure 4, also derived from Fig. 2a, is the value of the first derivative of the age distribution function and shows the velocity of the magmatic arc locus with respect to the trench as it swept inboard then back between 130 and 15 Myr. This velocity apparently attained almost 3 cm yr<sup>-1</sup> on the sweep in, collapsed to zero at the bend near 45 Myr, then increased to over 4 cm yr<sup>-1</sup> on the return. The duration of arc activity at any given place averages over 20 Myr during the sweep eastward and is somewhat less on the sweep back. The width of arc activity at any given time is initially narrow at steep angles of dip (about 100 km), but very wide (over 600 km) at low angles of dip or during periods of rapid migration.

## Discussion

If our interpretation that shifting magmatic arc patterns in southwestern North America are due to changes in dip of Cordilleran Benioff zones is correct, some explanation must be sought for the changing dips. Luyendyk<sup>24</sup> suggested that rates of convergence at subduction zones influenced Benioff zone dip with his formula: dip = arc sin  $V_z/V_0$ , where  $V_z$  is the natural rate a slab sinks due to gravity, and  $V_0$  is the rate of convergence. Thus, the higher the rate of convergence the lower the angle of dip. It has also been suggested that an actively driving continental plate, with a subduction zone dipping beneath the leading edge, pushes the subduction zone ahead, over-riding and entraining the Benioff zone below and decreasing the dip<sup>25</sup>. Other factors, such as age (temperature) and emboliment of aseismic ridges, might reduce slab density and influence Benioff zone dip<sup>26</sup>. Rates of convergence between an actively driving North America plate and Farallon plate may have increased during late Cretaceous–early Tertiary time, flattening the Benioff zone beneath western North America. Some time between 40–55 Myr the rate slowed, causing increase in Benioff zone dip and the resulting return sweep between 40 and 20 Myr.

Plate reconstructions based on assumed movement of Pacific and African plates over hotspots in the mantle since 135 Myr, although still suspect, yield vector subtraction derived convergent rates between North America and Farallon plates<sup>27,28</sup>. From 80 Myr to about 45 Myr the computed rates of convergence reach values in excess of 12 cm yr<sup>-1</sup>. The age of the bend in the Hawaii–Emperor seamount chain is placed at about 45 Myr (refs 29, 30) and is assumed to mark a change in direction of Pacific plate motion from northward to more north westward. After the change in direction, convergent rates between the two plates fall to less than 8 cm yr<sup>-1</sup>. We note that the age of the Hawaii–Emperor bend is the same as the age of the bend in Fig. 2a, and that high derived convergent rates correlate with the sweep inboard during Laramide time while the lower rates of post-45 Myr time correlate with the sweep back during mid-Tertiary time. Using Luyendyk's formula and a rate of convergence of 12 to 14 cm yr<sup>-1</sup>, the Benioff zone dip attains 20°. Slower rates of convergence between 6 and 8 cm yr<sup>-1</sup> yield dips from 40° to 60°. The high

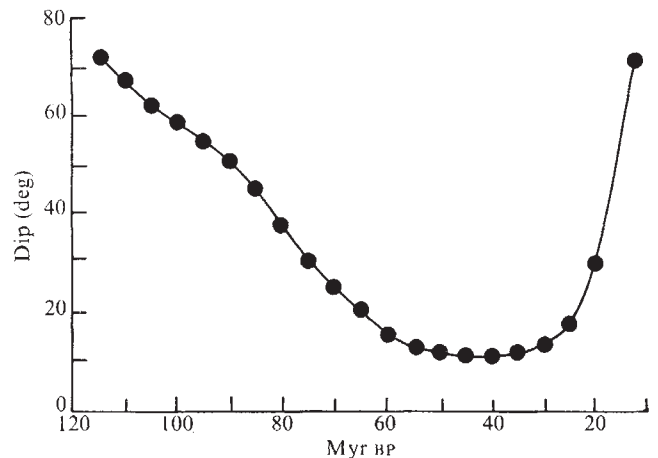


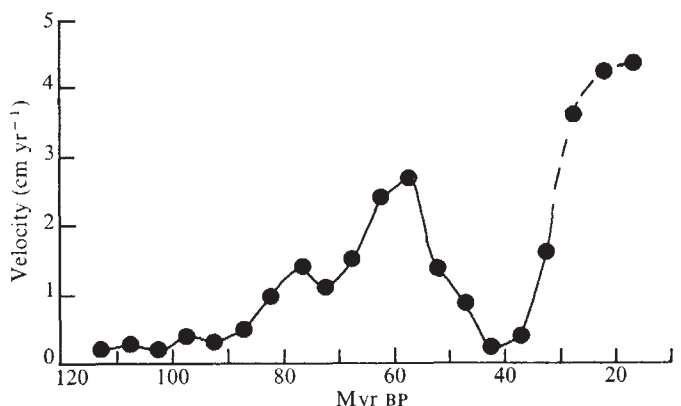
Fig. 3 Benioff zone dip angle as a function of time beneath southwestern North America. Graph derived from best-fit curve through data of Fig. 2a. A 15% post-15 Myr extension due to Basin and Range faulting has been removed from Fig. 2 to construct Fig. 3.

rates of convergence, coupled with a low-dipping Benioff zone during Laramide orogeny are thought to explain extensive deep-seated compressive thrusting which extended as far inboard from the trench as did the arc magmatism<sup>27,31</sup>.

The sweep in from 130 Myr to about 45 Myr includes two classic orogenic episodes in Cordilleran tectonics: the Sevier orogeny during mid-Cretaceous to about 80 Myr and the Laramide orogeny from 80 Myr to about 45 Myr (refs 9, 32, 33). The return sweep between 40 Myr and about 15 Myr includes what has been termed the mid-Tertiary orogeny<sup>34,35</sup>, or the ignimbrite flare-up<sup>28</sup>. In terms of magmatic arc activity, these 'orogenies' are here seen as a continuum broken only by the turn-around near 45 Myr which ended Laramide events and initiated the mid-Tertiary ignimbrite flare-up in southern Arizona. We note the latter part of the sweep inboard (Laramide time) produced a major percentage of North American copper porphyry deposits in southern Arizona, New Mexico, and Sonora. The return sweep across the same region produced essentially none.

We have preliminary evidence that the shifting magmatic arc patterns described for the southwestern Cordillera are related to similar patterns further north. Arc activity in the Sierra Nevada and Idaho batholiths began to disappear soon after 80 Myr (refs 19, 35). Just afterward a very sparse magmatic pulse swept eastward and rapidly died, producing the well

Fig. 4 Velocity of magmatic arc axis as a function of time. Graph derived from best-fit curve through data of Fig. 2a. A 15% post-15 Myr extension due to Basin and Range faulting has been removed from Fig. 2 to construct Fig. 4.



known Laramide 'igneous gap' between about 70 Myr and 55 Myr throughout the north-west and west-central United States<sup>36,10</sup>. At about 55 Myr the north-west erupted violently with wide-spread Challis-Absaroka volcanism and shallow plutonism<sup>36</sup>. This pattern then swept out of Washington, Idaho, Montana, and Wyoming, and moved progressively southwestward across western Utah and Nevada as the ignimbrite flare-up of the Great Basin between 40 and 20 Myr (refs 37–39). We interpret the sweep in and rapid extinguishing of arc activity as the result of the same flattening of a Benioff zone we describe to the south, the only difference being that in the north the dip became so shallow arc activity was impossible. The igneous gap extending from southern Ecuador to southern Peru, accompanied by an essentially horizontal Benioff zone<sup>40</sup>, is the present-day analogue. The Challis-Absaroka flare-up and subsequent return sweep of arc activity southwestward across the Great Basin would be the same return sweep we describe to the south and due to the same cause—a rapid increase in dip. Seen this way, the mid-Tertiary outburst of ignimbrites in the central and southern Cordillera becomes more comprehensible. Lipman *et al.*<sup>10</sup> called the outburst a magmatic arc related to subduction. We agree with this interpretation, and add that with the better resolution of time-space plots based on extensive radiometric dating the volcanism can be further seen as the rapid retrograde sweep of a magmatic arc toward the trench resulting from gravitational collapse and steepening of a Tertiary Benioff zone. Calling on complex back-arc diapirs<sup>41</sup> to explain mid-Tertiary volcanism in the Great Basin now seems unnecessary.

The data presented here suggest arc activity in southwestern North America was essentially continuous between 130 Myr and about 15 Myr. If a Kula-Farallon triple junction migrated up the North America plate margin during late Cretaceous-early Tertiary time, as suggested from some reconstructions<sup>42</sup>, we do not see evidence of its passage. The data seem more consistent with, but do not prove, continuous subduction of the Farallon plate during this time. Progressive destruction of the Pacific-Farallon spreading centre and Mendocino fracture zone along the North America plate margin during late Tertiary time<sup>43</sup> has bearing on interpretation of the youngest ages in the data array. Most reconstructions place this event between 30 Myr and 20 Myr, with initial contact somewhere along the coast of northernmost Baja California<sup>43</sup>. It is not certain to what degree the last part (25 Myr to 15 Myr) of the post-40 Myr collapse and rapid steepening of the Benioff zone was due to cessation of subduction caused by the change in plate geometry. Regardless, an arc activity delay-time of 5–10 Myr after slab truncation is predictable<sup>44</sup>.

The rapid retrograde migration of arc activity toward the trench after 40 Myr does not require excessive sinking rates

of the subducting slab. For example, between 40 Myr and 20 Myr the increase in dip-angle was about 20° (Fig. 3). This implies an approximate average vertical descent rate of about 1 cm yr<sup>-1</sup>. From 20 Myr to 15 Myr the increase in dip-angle was about 40° which yields an average sinking rate of about 5 cm yr<sup>-1</sup>. This 5 cm yr<sup>-1</sup> is the velocity Luyendyk<sup>24</sup> gives the natural vertical sinking rate ( $V_z$ ) of a slab due to gravity. This suggests the slab may actually have been detached and freely sinking after  $\pm 25$  Myr due to destruction of the East Pacific spreading centre and initiation of the San Andreas transform<sup>42</sup>.

Research was partially supported by a USNSF-Kennecott Exploration, Inc. Research Participation Grant. Paul Damon, Richard Nielson, and W. A. Rehrig provided unpublished data. We thank W. R. Dickinson, Paul Damon, and Warren Hamilton for helpful discussions.

Received 6 June; accepted 20 September 1977.

- McKenzie, D. P. & Parker, R. L. *Nature* **216**, 1276–1280 (1967).
- Isacks, B., Oliver, J. & Sykes, L. R. *J. geophys. Res.* **73**, 5855–5899 (1968).
- Dickinson, W. R. *Rev. Geophys.* **8**, 813–860 (1970).
- Karig, D. E. & Sharman, G. F. III *Geol. Soc. Am. Bull.* **86**, 377–389 (1975).
- Dickinson, W. R. *J. geophys. Res.* **78**, 3376–3389 (1973).
- Hatherton, T. & Dickinson, W. R. *J. geophys. Res.* **74**, 5301–5310 (1969).
- Isacks, B. & Molnar, P. *Rev. Geophys. Space Sci.* **9**, 103–174 (1971).
- Hamilton, W. *Geol. Soc. Am. Bull.* **80**, 2409–2430 (1969).
- Coney, P. J. *Nature* **233**, 462–465 (1971).
- Lipman, P. W., Prostka, J. H. & Christiansen, R. L. *Science* **174**, 821–825 (1971).
- Gilluly, J. *Geol. Soc. Am. Bull.* **82**, 2383–2396 (1971).
- Lowell, J. D. *Econ. Geol.* **69**, 601–617 (1974).
- Woodward, L. A. *Geology* **2**, 570 (1974).
- Gastil, R. G., Phillips, R. P. & Allison, E. C. *Geol. Soc. Am. Mem.* **140**, 170 (1975).
- Lanphere, M. N. & Reed, B. L. *Geol. Soc. Am. Bull.* **84**, 3773–3782 (1973).
- Hayes, P. T., Simons, F. S. & Raup, R. B. *U.S. Geol. Surv. Bull.* **1194-M**, 9 (1965).
- Silver, L. T., Early, T. O. & Anderson, T. H. *Geol. Soc. Am. Cordilleran Sec. Meet.* 375–376 (1975).
- Gastil, R. G. & Krummenacher, D. *Geol. Soc. Am. Bull.* **88**, 189–198 (1977).
- Armstrong, R. L. & Suppe, J. *Geol. Soc. Am. Bull.* **84**, 1375–1392 (1973).
- Krummenacher, D., Gastil, R. G., Bushee, J. & Doupoint, J. *Geol. Soc. Am. Bull.* **86**, 760–768 (1975).
- Anderson, T. H. & Silver, L. T. *Geol. Soc. Am., Rocky Mountain Meet.* 484 (1974).
- Damon, P. E. & Mauger, R. L. *Trans. S.M.E.* **235**, 99–112 (1966).
- Eberly, L. D. & Stanley, T. B. Jr. *Geol. Soc. Am. Bull.* (in the press).
- Luyendyk, B. P. *Geol. Soc. Am. Bull.* **81**, 3411–3416 (1970).
- Coney, P. J. in *Implications of Continental Drift to the Earth Sciences* (eds Tarling, D. H. & Runcorn, S. K.) **2**, 713–727 (Academic, London, 1973).
- Kelleher, J. & McCann, W. J. *geophys. Res.* **81**, 4885–4896 (1976).
- Coney, P. J. *New Mexico geol. Soc. Spec. Pub.* **4**, 5–10 (1976).
- Coney, P. J. *Geol. Soc. Am. Spec. Pap.* (in the press).
- Morgan, W. J. *et al. Geol. Soc. Am. Mem.* **132**, 7–22 (1972).
- Clague, D. A. & Jarrard, R. D. *Geol. Soc. Am. Bull.* **84**, 1135–1154 (1973).
- Burchfiel, B. C. & Davis, G. A. *Nature* **260**, 693–694 (1976).
- Armstrong, R. L. *Geol. Soc. Am. Bull.* **79**, 429–458 (1968).
- Heldrick, T. L. & Titley, S. R. *24th Int. Geol. Cong.* **3**, 740 (1976).
- Damon, P. E. *et al. A. Prog. Rep. A.E.C. COO-689-42* (1964).
- Armstrong, R. L., Taubeneck, W. H. & Hales, P. O. *Geol. Soc. Am. Bull.* **88**, 397–411 (1977).
- Armstrong, R. L. *Northwest Geol.* **3**, 1–15 (1974).
- Stewart, J. H., Moore, W. J. & Zeitz, I. *Geol. Soc. Am. Bull.* **88**, 67–77 (1977).
- Armstrong, R. L. & Higgins, R. F. *Geol. Soc. Am. Bull.* **84**, 1095–1100 (1973).
- Noble, D. C. *Earth planet. Sci. Lett.* **17**, 142–150 (1972).
- Stauder, W. J. *J. geophys. Res.* **80**, 1053–1064 (1975).
- Scholz, C. H., Barazangi, M. & Sbar, M. L. *Geol. Soc. Am. Bull.* **82**, 2979–2990 (1971).
- Atwater, T. *Geol. Soc. Am. Bull.* **81**, 3513–3536 (1970).
- Atwater, T. & Molnar, P. *Stanford Univ. Publ. Geol. Sci.* **13**, 136–149 (1973).
- Snyder, W. S., Dickinson, W. R. & Silberman, M. L. *Earth planet. Sci. Lett.* **32**, 91–106 (1976).
- Elston, W. E. & Northrop, S. A. *New Mexico Geol. Soc. Spec. Pub.* **5**, 151 (1976).

# Regulation of protein synthesis, intracellular electrolytes and cataract formation *in vitro*

Toshimichi Shinohara & Joram Piatigorsky

Section on Cellular Differentiation, Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20014

*Control of protein synthesis is associated with changes in the ratio of intracellular Na<sup>+</sup> to K<sup>+</sup> in the cultured embryonic chick lens. Correlations of intracellular Na<sup>+</sup>/K<sup>+</sup> ratios with crystallin synthesis and cataract formation in vitro suggest that the Na<sup>+</sup>/K<sup>+</sup> ratio may have an important role in the regulation of protein synthesis during cataractogenesis.*

CATARACTS, or lens opacities, represent a primary cause of seriously impaired vision and even blindness. At present, a cataract that has progressed beyond the initial stages cannot be reversed, forcing the patient to have the affected lens removed in order to restore clear vision. Cataracts are initiated by or associated with a variety of factors, including physical trauma, ageing, metabolic stress, hereditary defects, disease, ionising radiation and microwaves<sup>1</sup>. Although the



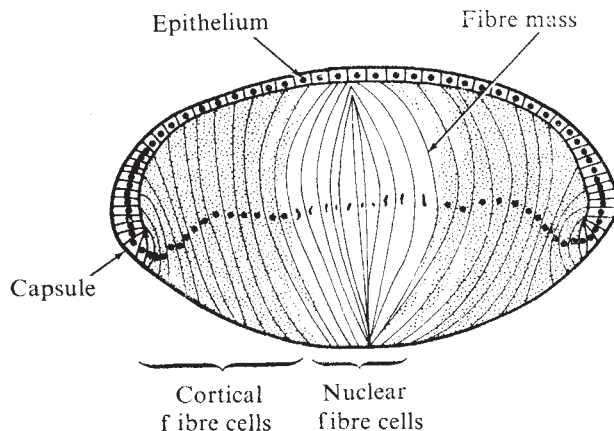
mechanism of cataractogenesis has not been definitively established for any one type of lens opacity, considerable evidence indicates that osmotic imbalance<sup>2</sup> and protein aggregation<sup>3-5</sup> may play prominent parts.

We have reported<sup>6</sup> that cortical cataracts are formed within 3 h after explantation of the embryonic chick lens into a defined culture medium (Ham's F-10<sup>7</sup>) (see Fig. 1). This experimental cataract seems to be caused by the trauma of detaching the vitreous body, a gelatinous connective tissue matrix. The initiation of this experimental cataract is associated with a specific alteration in the synthesis of  $\delta$ -crystallin, the principal protein of the embryonic chick lens<sup>8</sup>.  $\delta$ -Crystallin represents 60–80 % of the protein present in these lenses<sup>9-13</sup> and can be resolved into two bands with molecular weight near 50,000 and 48,000 by electrophoresis in a sodium dodecylsulphate-urea-polyacrylamide gel<sup>14</sup>. Clear lenses synthesise the proteins in these two bands in a ratio of approximately 1 : 3 in favour of the lower molecular weight band. In vitreous-free lenses developing cortical opacities the ratio of synthesis of proteins in these two bands is reversed to approximately 3 : 1 in favour of the higher molecular weight band. This alteration in protein synthesis is associated only with the initiation of the cataract, since these  $\delta$ -crystallin polypeptides are synthesised in the normal ratio after the lens has remained in culture for 24 h, despite persistence of the cortical opacity<sup>6</sup>.

Here we explore the possibility that the alteration of  $\delta$ -crystallin synthesis in the cultured lens is due to an increase in the intracellular concentration of Na<sup>+</sup> and a decrease in the intracellular concentration of K<sup>+</sup>, since several reports have shown that cataracts are associated with Na<sup>+</sup> influx and K<sup>+</sup> efflux<sup>2,15-20</sup>. In addition, one investigation has shown that stripping the vitreous body from a cultured rabbit lens accelerates cataract formation<sup>21</sup> and causes an increase in Na<sup>+</sup> and a decrease in K<sup>+</sup> concentration within the lens<sup>22</sup>. Moreover, experiments with bacterial<sup>23</sup> and mammalian<sup>24</sup> cells have indicated the importance of K<sup>+</sup> for the intracellular control of protein synthesis, and investigations of cell-free translation<sup>25</sup> have provided evidence that Na<sup>+</sup> and K<sup>+</sup> ions can discriminate between different mRNAs to allow preferential translation at different ionic strengths. Our data indicate the changes in the intracellular concentrations of Na<sup>+</sup> and K<sup>+</sup> cause the alteration of  $\delta$ -crystallin synthesis in the embryonic chick lens cultured without its vitreous body. This raises the possibility that the intracellular levels of these ions also contribute to the control of differential protein synthesis in other eukaryotic cells; this in turn may influence cellular growth, differentiation and pathogenesis.

### Na<sup>+</sup> and K<sup>+</sup> in cultured lenses with and without vitreous body

Na<sup>+</sup> and K<sup>+</sup> concentrations were determined in the embryonic chick lens immediately after explantation, after 3 h of culture with its vitreous body attached, and after 3 or 24 h of culture with its vitreous body detached (Table 1). At the time of explantation, the concentrations of these ions were similar to those found for lenses of other species<sup>26</sup>, with Na<sup>+</sup> being considerably lower than K<sup>+</sup>. The Na<sup>+</sup> and K<sup>+</sup> concentrations remained the same after the lenses were cultured for 3 h with their vitreous bodies; these lenses with attached vitreous body have been shown to be transparent and synthesise the proteins comprising the two bands of  $\delta$ -crystallin in the normal ratio of 1 : 3 in favour of the lower molecular weight band<sup>6</sup>. By contrast, after removal of the vitreous body, the Na<sup>+</sup> concentration increased approximately sixfold and the K<sup>+</sup> concentration decreased approximately 2.5-fold within 3 h; these vitreous-free lenses have been shown to possess cortical cataracts and synthesise the proteins of the two  $\delta$ -crystallin bands in a ratio of approximately 3 : 1 in favour of the higher molecular weight band<sup>6</sup>.



**Fig. 1** Diagrammatic representation of a cross section of a 15-d-old embryonic chick lens. The nuclei in the centre of the fibre mass (nuclear fibre cells) are starting to become pycnotic and will eventually disintegrate. The stippled zone represents the area which becomes opaque when the lens is cultured after its vitreous body has been removed from the posterior surface. The vitreous body (not shown) occupies a major part of the eye and consists of an avascular, viscous connective tissue matrix containing fibroblasts, collagenous material and other proteins, mucopolysaccharides and organic substances.

After 24 h *in vitro* without the vitreous body, the Na<sup>+</sup> concentration had returned to nearly its initial value at explantation; the vitreous-free lens after 24 h *in vitro* has been shown to still have the cortical cataract, but to return to a relatively normal pattern of  $\delta$ -crystallin synthesis<sup>6</sup>. Since the intracellular K<sup>+</sup>/Na<sup>+</sup> ratio and pattern of  $\delta$ -crystallin synthesis are markedly different in the lenses cultured with or without the vitreous body, we were encouraged to pursue the relationship between the intracellular concentrations of Na<sup>+</sup> and K<sup>+</sup> and the ratio of synthesis of the  $\delta$ -crystallin polypeptides.

### Effect of ouabain on protein synthesis in cultured lenses

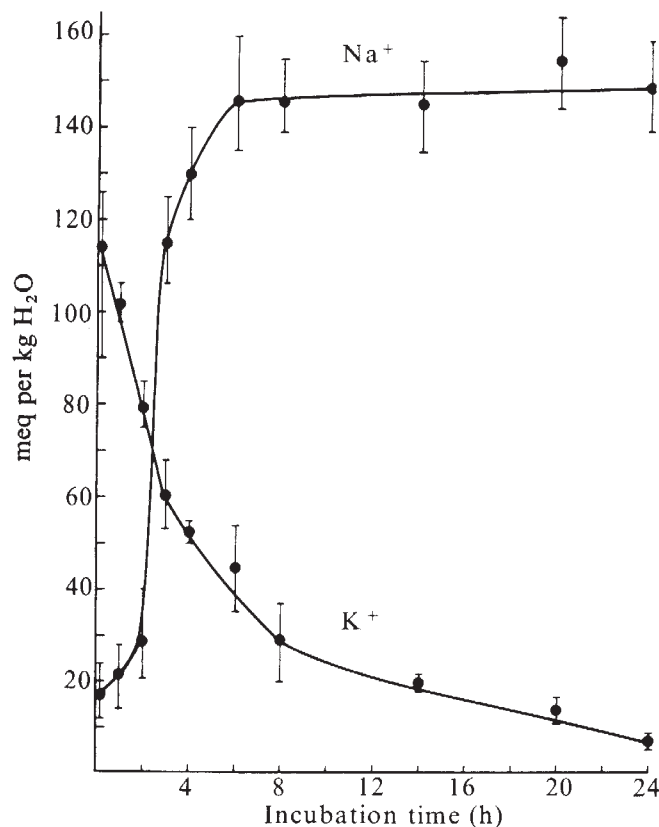
In order to alter the intralenticular concentrations of Na<sup>+</sup> and K<sup>+</sup> without removing the vitreous body, explanted lens-vitreous preparations were treated with ouabain, an inhibitor of Na<sup>+</sup>,K<sup>+</sup> ATPase activity<sup>27</sup>. This enzyme is responsible for the permeability of Na<sup>+</sup> and K<sup>+</sup> in eukaryotic cells. Inhibition of its activity with ouabain is known

**Table 1** K<sup>+</sup> and Na<sup>+</sup> concentrations in embryonic chick lenses

|                                | K <sup>+</sup> | mEq per kg H <sub>2</sub> O<br>Na <sup>+</sup> | K <sup>+</sup> /Na <sup>+</sup> |
|--------------------------------|----------------|------------------------------------------------|---------------------------------|
| Immediately after explantation | 112 ± 11       | 17 ± 3                                         | 6.6                             |
| Cultured 3 h, + vitreous body  | 110 ± 7        | 16 ± 6                                         | 6.9                             |
| Cultured 3 h, - vitreous body  | 40 ± 3         | 84 ± 9                                         | 0.5                             |
| Cultured 24 h, - vitreous body | 47 ± 8         | 29 ± 18                                        | 1.6                             |

Four to eight lenses from 15-d-old chick embryos with or without their vitreous body were cultured in plastic tissue culture dishes (Falcon, 60 × 15 mm) containing 10 ml of Ham's F-10 medium supplemented with fructose (848 mg per 100 ml); all assays were performed on lenses stripped of their vitreous body. Each determination represents an average of five to eight experiments. The lenses were washed quickly in groups of four with deionised water, placed in pyrex test tubes (previously soaked overnight in concentrated nitric acid to remove adhering Na<sup>+</sup>), dried at 75 °C for 24 h, dissolved in 100  $\mu$ l of concentrated nitric acid for several hours at room temperature and diluted with 40 ml of deionised water. K<sup>+</sup> was determined by flame photometry and Na<sup>+</sup> by atomic absorption spectrometry (Perkin-Elmer model 603). The average water content of the embryonic lenses was 85 % of their total weight. Corrections were made for ions in the extracellular space of the lens, which was determined by experiments with <sup>3</sup>H-inulin.





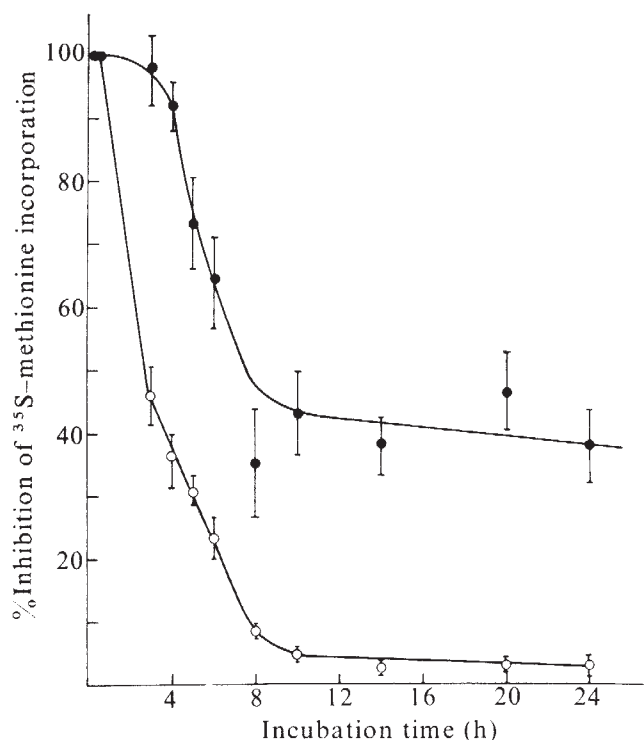
**Fig. 2** Effect of ouabain on the concentrations of Na<sup>+</sup> and K<sup>+</sup> in cultured embryonic chick lenses. Each point on the graph represents an average of three experiments; the vertical lines indicate the range of values obtained in the three tests. The lenses were cultured with their vitreous body and analysed for Na<sup>+</sup> and K<sup>+</sup> as given in Table 1.

to increase the Na<sup>+</sup> and decrease the K<sup>+</sup> concentration in cultured lenses of other species<sup>28-31</sup>. Ouabain caused a reciprocal change in the relative concentrations of the ions of the order of tenfold for each ion in the cultured lenses (Fig. 2). The ouabain-treated embryonic lens-vitreous preparations developed a diffuse opacity covering the entire epithelium and the most anterior aspect of the fibre mass. The bulk of the fibres remained clear, however, even after 24 h *in vitro*, and did not develop cortical cataracts unless the vitreous body was detached.

Incorporation of <sup>35</sup>S-methionine into protein was severely affected by ouabain treatment. Lenses labelled for 3 h with <sup>35</sup>S-methionine incorporated about 90% less isotope into their total protein after 24 h of preincubation with  $1 \times 10^{-4}$  M ouabain than after a comparable preincubation in the absence of the drug. The pattern of incorporation into  $\delta$ -crystallin was strikingly altered in lenses treated with ouabain. Incorporation of <sup>35</sup>S-methionine into the protein of the lower molecular weight  $\delta$ -crystallin band was inhibited by at least 95% after 8 h of culture with ouabain, while incorporation of the <sup>35</sup>S-methionine into the protein of the higher molecular band was only inhibited 50–60% (Fig. 3). This greater inhibition of methionine incorporation into the protein of the lower molecular weight band of  $\delta$ -crystallin persisted as long as ouabain was present in the culture medium. The staining patterns of the polyacrylamide gels, reflecting the absolute amount of protein present in each band, were similar in the preparations from lenses cultured with or without ouabain. The difference in <sup>35</sup>S-methionine incorporation into the proteins of the two  $\delta$ -crystallin bands was apparently not due to differential degradation of the newly synthesised polypeptides, since the protein in neither band was degraded in lenses labelled with <sup>35</sup>S-methionine

for 3 h in the absence of ouabain and subsequently incubated for 24 h in the presence of ouabain and the absence of protein synthesis (inhibited by  $10 \mu\text{g ml}^{-1}$  cycloheximide). Since the rate of the total uptake of <sup>35</sup>S-methionine into the lens did not decrease over the 24 h of culture (data not shown), the differential decrease of <sup>35</sup>S-methionine incorporation into protein seems to be due to differential inhibition of synthesis of the  $\delta$ -crystallin polypeptides.

It is important to note that the change in the ratio of synthesis of the  $\delta$ -crystallin polypeptides occurred in the epithelium and the fibre mass of the ouabain-treated lens, but was confined to the fibre mass in the vitreous-free lens. We believe that this difference in the localisation of the alteration of  $\delta$ -crystallin synthesis is due to a corresponding difference in the localisation of the changes in the levels of Na<sup>+</sup> and K<sup>+</sup>. The Na<sup>+</sup>,K<sup>+</sup> ATPase activity of the lens is located principally<sup>29,32,33</sup>, although apparently not exclusively<sup>34,35</sup>, in the epithelium. The ionic balance of the lens is thus thought to be governed by a pump-leak system<sup>36</sup>. Na<sup>+</sup> diffuses into the lens through the fibre mass at the posterior surface and is transported out of the lens through the epithelium at the anterior surface; by contrast, K<sup>+</sup> is transported into the lens through the epithelium and diffuses out of the lens through the fibre mass. Thus, detaching the vitreous body from the posterior surface may increase Na<sup>+</sup>



**Fig. 3** Effect of  $1 \times 10^{-4}$  M ouabain on the relative incorporation of <sup>35</sup>S-methionine into the proteins of the higher and lower molecular weight bands of  $\delta$ -crystallin. The lenses were cultured with their vitreous body as for Table 1. Each point on the graph represents an average of three experiments; the vertical lines indicate the range of values obtained in the three tests. Lenses were labelled individually for 3 h in 1 ml of medium containing  $250 \mu\text{Ci}$  of <sup>35</sup>S-methionine (New England Nuclear,  $450 \text{ Ci mmol}^{-1}$ ). One to  $2 \mu\text{g}$  of total homogenate protein was examined by electrophoresis in a discontinuous, 10% polyacrylamide gel containing 0.1% sodium dodecylsulphate and 8 M urea as described elsewhere<sup>11</sup>. The stained  $\delta$ -crystallin bands were cut from the gel, dissolved in  $100 \mu\text{l}$  of hydrogen peroxide and assayed for radioactivity by scintillation counting. The % inhibition of incorporation by ouabain into the protein of each band was calculated by dividing the amount of radioactivity in that band at each time point by the amount of radioactivity in the same band from lenses labelled for 3 h without ouabain immediately after explantation. ●, Higher molecular weight band; ○, lower molecular weight band.

diffusion into the lens without significantly changing the concentrations of  $\text{Na}^+$  and  $\text{K}^+$  in the epithelium. Ouabain treatment, however, should increase the  $\text{Na}^+$  concentration and decrease the  $\text{K}^+$  concentration in both the fibre mass and the epithelium by inhibiting the  $\text{Na}^+, \text{K}^+$  ATPase activity of the lens.

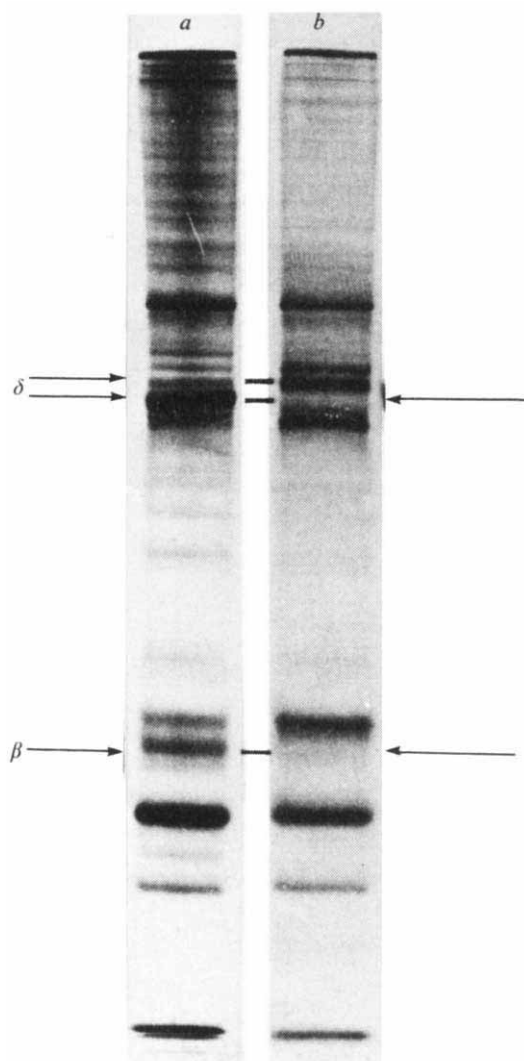
In addition to the alteration of  $\delta$ -crystallin synthesis, ouabain treatment also caused the differential inhibition of synthesis of other proteins in the cultured lenses. This was best observed when examining the isolated epithelium where  $\delta$ -crystallin represents a much smaller proportion of the total protein than in the fibre mass. The autoradiograms shown in Fig. 4 demonstrate the marked inhibition of synthesis of a  $\beta$ -crystallin component as well as that of the protein in the lower molecular weight  $\delta$ -crystallin band in lenses treated with ouabain. The synthesis of this  $\beta$ -crystallin polypeptide is also inhibited in lenses developing cortical cataracts in culture after their vitreous body has been removed. Inspection of the autoradiograms of Fig. 4 reveals a number of other differences in the relative amount of  $^{35}\text{S}$ -methionine incorporated into different bands of protein. Since equal amounts of radioactivity were applied to the gels, the synthesis of some bands of protein seems to be selectively stimulated by ouabain, but these actually represent proteins whose synthesis is not appreciably inhibited by the agent.

### Regulation of protein synthesis by $\text{Na}^+$ and $\text{K}^+$

To test the effect of  $\text{Na}^+$  and  $\text{K}^+$  on protein synthesis we examined cultured lenses with different intracellular concentrations of these ions. This was made possible by varying the amounts of  $\text{Na}^+$  and  $\text{K}^+$  in the culture medium of the lenses, which were permeable to these cations after treatment with ouabain. Determination of the intralenticular  $\text{Na}^+$  and  $\text{K}^+$  concentrations confirmed that the ouabain-treated lens was permeable to  $\text{Na}^+$  and  $\text{K}^+$ , and that these ions were at a similar concentration within the lens and in the medium. Tests with  $^3\text{H}$ -inulin, a compound which does not penetrate into cells, established that the ions were principally in the lens cells rather than in the extracellular space.

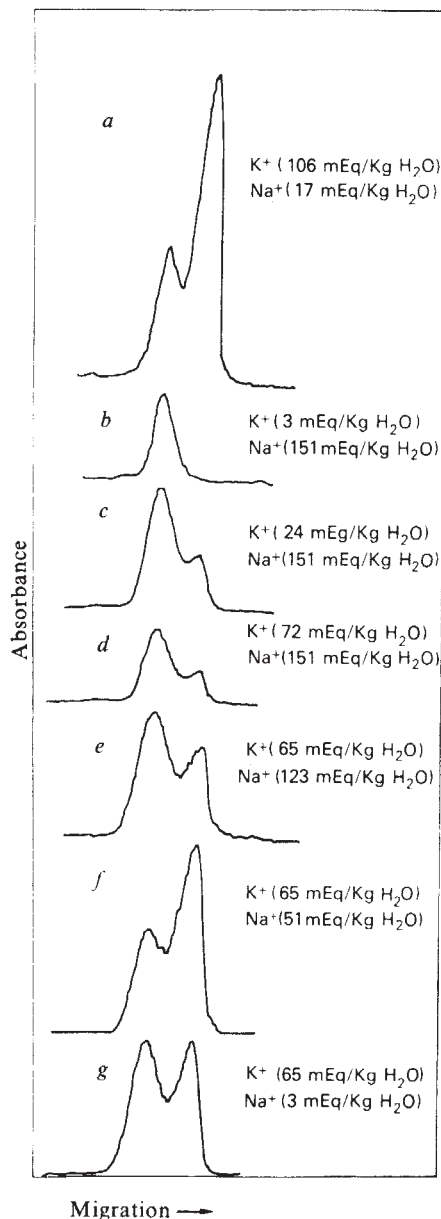
The incorporation of  $^{35}\text{S}$ -methionine into protein was sensitive to changes in the total intracellular concentration of  $\text{Na}^+$  and  $\text{K}^+$ .  $^{35}\text{S}$ -methionine incorporation was severely inhibited at combined  $\text{Na}^+$  and  $\text{K}^+$  concentrations above 200 mM or below 10 mM.  $\text{K}^+$  concentrations above 96 mM almost completely inhibited  $^{35}\text{S}$ -methionine incorporation into protein. These inhibitions of  $^{35}\text{S}$ -methionine incorporation into protein at different ionic strengths were apparently due to decreases in protein synthesis rather than  $^{35}\text{S}$ -methionine uptake since the total accumulation of radioactivity was not appreciably different in lenses cultured in the presence of ouabain in medium containing different concentrations of  $\text{Na}^+$  and  $\text{K}^+$ .

The most striking result of these tests was the sensitivity of the ratio of synthesis of the  $\delta$ -crystallin polypeptides to the variations in the concentrations of  $\text{Na}^+$  and  $\text{K}^+$ . The differential inhibition of synthesis of the protein in the lower molecular weight band of  $\delta$ -crystallin induced by ouabain (Fig. 5b) was partially overcome by increasing the concentration of  $\text{K}^+$  (Fig. 5c, d). Lowering the  $\text{Na}^+$  concentration also increased the relative amount of  $^{35}\text{S}$ -methionine incorporated into the protein in the lower molecular weight band of  $\delta$ -crystallin (Fig. 5e, f). The differential synthesis of the  $\delta$ -crystallin polypeptides does not seem to be simply a function of the total intracellular concentration of  $\text{Na}^+$  plus  $\text{K}^+$  because the protein in the lower molecular weight band was not synthesised at 154 mEq per kg of water of  $\text{Na}^+$  plus  $\text{K}^+$  (Fig. 5b) but was synthesised at higher (Fig. 5c–e) and lower (Fig. 5f, g) concentrations of intracellular  $\text{Na}^+$  plus  $\text{K}^+$ . Furthermore, the combined concentration of  $\text{Na}^+$  plus  $\text{K}^+$  in lenses cultured



**Fig. 4** Autoradiograms demonstrating the effect of ouabain on protein synthesis in the epithelium of the embryonic chick lens. Lenses were cultured in the absence (a) or presence (b) of  $1 \times 10^{-4}$  M ouabain for 24 h and then labelled with  $^{35}\text{S}$ -methionine for 3 h as given in Fig. 3. The epithelium was separated from the fibre mass and the proteins subjected to discontinuous polyacrylamide gel electrophoresis in the presence of sodium dodecylsulphate and urea<sup>11</sup>. The gels were stained and autoradiographed. The arrows on the right denote the almost complete inhibition of  $^{35}\text{S}$ -methionine incorporation into the proteins in the lower molecular weight  $\delta$ -crystallin band and into one  $\beta$ -crystallin band by ouabain. The staining patterns of the two samples were the same and were similar to the autoradiographic pattern of (a).

for 3 h with or without the vitreous body was the same, but the ratio of synthesis of the  $\delta$ -crystallin polypeptides has been shown to be different in the two cases<sup>6</sup> and varies with the ratio of the concentration of these ions (see Table 1). These results were not due to a differential processing of the higher molecular weight  $\delta$ -crystallin protein to a lower molecular weight form at different concentrations of  $\text{Na}^+$  and  $\text{K}^+$ , because there was no transfer of radioactivity from the higher to the lower molecular weight band in preparations from lenses labelled for 3 h in the presence of ouabain in the ionic conditions described in Fig. 5b and subsequently incubated for 2 d with ouabain in the absence of protein synthesis (inhibited by  $10 \mu\text{g ml}^{-1}$  cycloheximide) in the ionic conditions given in Fig. 5f. The differential inhibition of synthesis of the non- $\delta$ -crystallin proteins was also eliminated by adjusting the concentration of  $\text{Na}^+$  and  $\text{K}^+$  in the lenses treated with ouabain (data not shown). Thus, intracellular  $\text{K}^+/\text{Na}^+$  ratios greater than 1 generally allowed



**Fig. 5** Scans of autoradiograms demonstrating the effects of  $\text{Na}^+$  and  $\text{K}^+$  on the synthesis of the higher and lower molecular weight polypeptides of  $\beta$ -crystallin. Embryonic chick lenses were cultured for 24 h in the absence (a) or presence (b–g) of  $1 \times 10^{-4}$  M ouabain and labelled with  $^{35}\text{S}$ -methionine for 3 h as in Fig. 3, except that the  $\text{Na}^+$  and  $\text{K}^+$  concentrations in the medium were varied. Approximately equal amounts of  $\delta$ -crystallin (1–2  $\mu\text{g}$ ) were examined by electrophoresis and autoradiography, and the  $\delta$ -crystallin bands were scanned with a Quick Scan Jr densitometer (Helena Laboratories). The  $\text{Na}^+$  and  $\text{K}^+$  values were (in mEq per kg  $\text{H}_2\text{O}$ ): a,  $\text{K}^+$ , 106;  $\text{Na}^+$ , 17; b,  $\text{K}^+$ , 3;  $\text{Na}^+$ , 151; c,  $\text{K}^+$ , 24;  $\text{Na}^+$ , 151; d,  $\text{K}^+$ , 72;  $\text{Na}^+$ , 151; e,  $\text{K}^+$ , 65;  $\text{Na}^+$ , 123; f,  $\text{K}^+$ , 65;  $\text{Na}^+$ , 51; g,  $\text{K}^+$ , 65;  $\text{Na}^+$ , 3. These values represent intracellular levels of these ions which, except for (a), were similar to the concentrations in the medium, since ouabain made the lens permeable to  $\text{Na}^+$  and  $\text{K}^+$ . The concentrations of  $\text{Na}^+$  and  $\text{K}^+$  in Ham's F-10 medium (a) are 151 mEq per kg water and 3 mEq per kg water, respectively. Deficiencies in  $\text{Na}^+$  were compensated by adding choline chloride to the medium (e–g).

the normal ratio of synthesis of the proteins in the two  $\delta$ -crystallin bands, while  $\text{K}^+/\text{Na}^+$  ratios less than 1 favoured the synthesis of the protein in the higher molecular weight band of  $\delta$ -crystallin (see Table 1 also). The only deviation from this behaviour that we have observed is presented in Fig. 5g, where an extremely low  $\text{Na}^+$  concentration permitted more synthesis of the protein in the lower molecular

weight  $\delta$ -crystallin band than would have occurred at a slightly higher concentration of  $\text{Na}^+$ . The alteration of  $\delta$ -crystallin synthesis, then, is not entirely a simple function of the intracellular  $\text{Na}^+/\text{K}^+$  ratio.

## Discussion

These present data have implications with respect to the biology and pathology of the lens and the control of protein synthesis in general at the cellular and molecular levels. Experiments with cultured rabbit lenses showed that a disturbance of the normal lens-vitreous relationship alters the normal intralenticular  $\text{Na}^+$  and  $\text{K}^+$  concentrations<sup>22</sup>. Our experiments extend this observation to the embryonic chick lens and in addition demonstrate that the altered intracellular  $\text{Na}^+$  and  $\text{K}^+$  levels are correlated with differential changes in protein synthesis. It is important to note, however, that we do not know if the differential inhibition of protein synthesis is controlled directly or indirectly by the intracellular levels of  $\text{Na}^+$  and  $\text{K}^+$ . The  $\text{Na}^+$  and  $\text{K}^+$  concentrations of the lens are known to change during cataractogenesis in other species<sup>2,15–20</sup>, so it is possible that there are associated disturbances in protein synthesis which, except for the cultured chick lens<sup>6</sup>, have not yet been observed. It is not known whether such putative changes in protein synthesis might have a causal role in the cataractogenic process in some cases. The cortical cataract formed in the embryonic chick lens cultured without its vitreous body is not due to the alteration in  $\delta$ -crystallin synthesis because the same disturbance in synthesis takes place in lenses treated with ouabain but the cortical fibres remain clear. The lens epithelia, however, become cloudy in the ouabain-treated lenses.

Deviations from the normal pattern of polypeptide synthesis, such as demonstrated here, may have consequences extending to higher levels of protein structure. Lens crystallins may be particularly susceptible to changes in their native structure since they are aggregates of polypeptides<sup>8,37,38</sup>. The alteration in synthesis of the ratio of the  $\delta$ -crystallin polypeptides does indeed result in new forms of the native protein, which is composed of four subunits<sup>39</sup>. These new species of  $\delta$ -crystallin, to be reported elsewhere, are still tetramers but consist of different combinations of  $\delta$ -crystallin subunits. Clearly alterations in the structure of native proteins may also affect their function.

Our earlier experiments involving cell-free translation of purified  $\delta$ -crystallin mRNA suggest that the control of  $\delta$ -crystallin synthesis by  $\text{Na}^+$  and  $\text{K}^+$  is mediated at the translational level<sup>14</sup>. Nonetheless, we have not excluded the possibility that one of the  $\delta$ -crystallin bands is post-translationally derived from the other. This would be consistent with the observation that the two bands have similar (although not necessarily identical) tryptic peptides<sup>14,40</sup>. A post-translational modification of either band into the other would have to be associated with translation, since the bands do not interconvert after they have been synthesised<sup>14</sup>. Differential translation has been reported for immunoglobulin mRNAs<sup>41</sup>, HeLa cell mRNAs<sup>42</sup> and viral mRNAs<sup>43–45</sup> in cells under suboptimal conditions for protein synthesis. Evaluation of the literature of many different systems<sup>25,46–49</sup> indicates that translational control of protein synthesis generally occurs at the level of chain initiation. Thus, if the present control of  $\delta$ -crystallin synthesis by  $\text{Na}^+$  and  $\text{K}^+$  functions at the translational level, it is likely that this involves regulation of initiation. This would represent a very interesting example of translational control of protein synthesis in view of the apparent homogeneity of  $\delta$ -crystallin mRNA<sup>39</sup>.

It has been pointed out previously that the intracellular concentrations of  $\text{Na}^+$  and  $\text{K}^+$  may affect the absolute amount of macromolecular synthesis, and thus perhaps growth, in eukaryotic cells<sup>24</sup>. In view of the present data,



one may further speculate that the control of specific protein synthesis during differentiation may be affected by differences in the intracellular levels of  $\text{Na}^+$  and  $\text{K}^+$ . For example, the relative proportions of crystallins synthesised change during differentiation of the lens and differ in different compartments of the lens<sup>8,10-13,51</sup>. Possibly gradients of ions within the lens play a part in this regulation of crystallin synthesis.

We thank Mr L. Merola for use of a flame photometer and advice on the determination of  $\text{Na}^+$  and  $\text{K}^+$ , Drs B. Chock and J. Jones for use of an atomic absorption spectrometer, Ms C. Kunkle and C. Harvey for typing the manuscript, and Drs J. H. Kinoshita, P. Zelenka, D. Beebe, R. Crouch and A. J. Coulombre for useful comments.

Received 11 August; accepted 26 September 1977.

1. *Cataract and Abnormalities of the Lens* (ed. Bellows, J. G.) (Grune and Stratton, New York, 1975).
2. Kinoshita, J. H. *Invest. Ophthalmol.* **13**, 713-724 (1974).
3. Benedek, G. B. *Appl. Opt.* **10**, 459-473 (1971).
4. Spector, A. *Isr. J. med. Sci.* **8**, 1577-1582 (1972).
5. Jedziniak, J. A., Kinoshita, J. H., Yates, E. M., Hocker, L. O. & Benedek, G. B. *Invest. Ophthalmol.* **11**, 905-915 (1972).
6. Piatigorsky, J. & Shinohara, T. *Science* **196**, 1345-1347 (1977).
7. Ham, R. G. *Expl Cell Res.* **29**, 515-526 (1963).
8. Clayton, R. M. in *The Eye* vol. 5 (eds Davson, H. & Graham, L. T.) 400-494 (Academic, London, 1974).
9. Rabaey, M. *Expl Eye Res.* **1**, 310-316 (1962).
10. Zwaan, J. & Ikeda, A. *Expl Eye Res.* **7**, 301-311 (1968).
11. Genis-Galvez, J. M., Maisel, H. & Castro, J. M. *Expl Eye Res.* **7**, 593-602 (1968).
12. Yoshida, K. & Katoh, A. *Expl Eye Res.* **11**, 184-194 (1971).
13. Piatigorsky, J., Webster, H. deF. & Craig, S. P. *Dev Biol.* **27**, 176-189 (1972).
14. Reszelbach, R., Shinohara, T. & Piatigorsky, J. *Expl Eye Res.* (in the press).

15. Chylack, L. T., Jr & Kinoshita, J. H. *Invest. Ophthalmol.* **8**, 401-412 (1969).
16. Andr  e, G. *Ber. Versamm. dt. ophthalm. Ges.* **70**, 354-358 (1970).
17. Iwata, S. & Kinoshita, J. H. *Invest. Ophthalmol.* **10**, 504-512 (1971).
18. van Heyningen, R. *Expl Eye Res.* **13**, 136-147 (1972).
19. Maraini, G. & Mangili, R. *Ciba Fdn Symp.* **19** (new ser.), 79-97 (1973).
20. Duncan, G. & Bushell, A. R. *Expl Eye Res.* **20**, 223-230 (1975).
21. Chylack, L. T., Jr & Kinoshita, J. H. *Expl Eye Res.* **14**, 58-64 (1972).
22. Chylack, L. T., Jr & Kinoshita, J. H. *Expl Eye Res.* **15**, 61-69 (1973).
23. Lubin, M. & Ennis, H. L. *Biochim. biophys. Acta* **80**, 614-631 (1964).
24. Lubin, M. *Nature* **213**, 451-453 (1967).
25. Carrasco, L. & Smith, A. E. *Nature* **264**, 807-809 (1976).
26. Kuck, J. F. R., Jr. in *Biochemistry of the Eye* (ed. Graymore, C. N.) 183-260 (Academic, London, 1970).
27. Dahl, J. L. & Hokin, L. E. *A. Rev. Biochem.* **43**, 327-356 (1974).
28. Kinoshita, J. H., Kern, H. L. & Merola, L. O. *Biochim. biophys. Acta* **47**, 458-466 (1961).
29. Bonting, S. L., Caravaggio, L. L. & Hawkins, N. M. *Archs Biochem. Biophys.* **101**, 47-55 (1963).
30. Bonting, S. L. *Invest. Ophthalmol.* **4**, 723-738 (1965).
31. Harris, J. E. & Becker, B. *Invest. Ophthalmol.* **4**, 709-722 (1965).
32. Becker, B. & Cotlier, E. *Invest. Ophthalmol.* **1**, 642-645 (1962).
33. Kinsey, V. E. & Reddy, D. V. N. *Invest. Ophthalmol.* **4**, 104-116 (1965).
34. Riley, M. V. *Expl Eye Res.* **9**, 28-37 (1970).
35. Duncan, G. *Ciba Fdn Symp.* **19** (new ser.), 99-116 (1973).
36. Kinsey, V. E. *Eye Structure. II Symp.* (ed. Rohen, J. W.) 383-394 (Schattauer, Stuttgart, 1965).
37. Spector, A. *Invest. Ophthalmol.* **4**, 579-591 (1965).
38. Bloemendal, H. *Science* **197**, 127-138 (1977).
39. Piatigorsky, J., Zelenka, P. & Simpson, R. T. *Expl Eye Res.* **18**, 435-446 (1974).
40. Piatigorsky, J. *J. biol. Chem.* **251**, 4416-4420 (1976).
41. Sonenshein, G. E. & Brawerman, G. *Biochemistry* **15**, 5497-5501 (1976).
42. Saborio, J. L., Pong, Sheng-Shung & Koch, G. *J. molec. Biol.* **85**, 195-211 (1974).
43. Nuss, D. L., Oppermann, H. & Koch, G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1258-1262 (1975).
44. Nuss, D. L. & Koch, G. *J. Virol.* **17**, 283-286 (1976).
45. Schochetman, G. & Schlom, J. *Virology* **73**, 431-441 (1976).
46. Lodish, H. *J. biol. Chem.* **246**, 7131-7138 (1971).
47. McKeehan, W. L. *J. biol. Chem.* **249**, 6517-6526 (1974).
48. Sonenshein, G. E. & Brawerman, G. *Biochemistry* **15**, 5501-5506 (1976).
49. Lodish, H. *A. Rev. Biochem.* **45**, 39-72 (1976).
50. Zelenka, P. & Piatigorsky, J. *Expl Eye Res.* **22**, 115-124 (1976).
51. Delcour, J. & Papaconstantinou, J. *Biochem. biophys. Res. Commun.* **57**, 134-141 (1974).

# letters to nature

## Quasar-galaxy pairs and surface density of quasars

DISCOVERIES of radio-quiet quasars in the vicinity of galaxies<sup>1-5</sup>, sometimes with possible signs of interactions seem to contradict strongly the statistical analyses<sup>6-9</sup> on quasar-galaxy associations, which do not support the previous study<sup>10</sup> involving 3C quasars and bright galaxies. In each of these analyses<sup>6-10</sup>, the calculations used the surface densities of galaxies, because the surface density of quasars is poorly known, but it has been claimed<sup>11</sup> that the small angular separations are explicable by chance only if the surface density of quasars is much larger than believed. To test such an assumption we studied the problem by estimating directly the probabilities involved when adopting the most probable values of the density of quasars<sup>12,13</sup> at  $B \sim 19.0-19.5$ , that is,  $\mu \sim 3-5 \text{ (deg)}^{-2}$ .

The formula  $f_1(r) = 2\pi\mu r \exp(-\pi\mu r^2)$  gives the distribution of the angular separations ( $r$ ) between any galaxy and its nearest quasar. The characteristics of this distribution for different values of  $\mu$  are shown in Table 1 with the lower limit of the bilateral 95% confidence level range (column 4). The curves are shown in Fig. 1 with the lower limit taken from column 4 of table 1. For  $\mu \sim 3-5 \text{ (deg)}^{-2}$ , the curves suggest that quasars 3 arc min or more from a galaxy should not be called 'quasars near galaxies' as they occur by chance with more than 2.5% probability. Table 2 lists every radio-quiet quasar within 3 arc min of a galaxy (Arp, ref. 11) (MKN205 is excluded as a compact galaxy<sup>14</sup>). If we assume that Arp's search technique<sup>11</sup> yields quasars of the same nature as the other quasars (there is no evidence to the contrary<sup>15</sup>), for  $\mu = 3$ , nine such pairs would have been found if a search had been made using  $N = 386$  randomly selected quasars ( $P = 2.33\%$ ) and for  $\mu = 5$ ,  $N = 237$  ( $P = 3.80\%$ ). There is then no strong discrepancy with the presently known 140 radio-quiet quasars in the last

Reference Catalogue of Quasars<sup>16</sup>, since only two of these nine were not specifically searched for in the vicinity of galaxies. Moreover, 2 is less than the 3.3 and 5.3 expected pairs for  $\mu = 3$  and 5 respectively with  $N = 140$ .

Some other important effects are (1) two objects in Table 2 are not yet confirmed as quasars.

(2) We need a limiting magnitude for the galaxy, because, for every quasar, we could find a close foreground (or background) galaxy, however faint this galaxy is, if we were not limited by the sky brightness. Then we should exclude cases where very faint galaxies are involved, such as the 18 mag galaxy near 0846+51. Moreover, if quasars are associated with bright galaxies, as already suggested, how can they also be found at the distances of the faintest galaxies?

(3) Homogeneous samples of radio-quiet quasars<sup>7</sup> as well as the quasars recently<sup>17,18</sup> detected do not depart from a random distribution with respect to galaxies.

**Table 1** Distribution of the angular distances between quasars and galaxies if quasars are randomly distributed with a mean density  $\mu$

| $\mu$<br>(deg) <sup>-2</sup> | Mean<br>value<br>(arc min) | Variance<br>(arc min) <sup>2</sup> | Lower<br>95% limit<br>(arc min) |
|------------------------------|----------------------------|------------------------------------|---------------------------------|
| 1                            | 29.9                       | 244.0                              | 5.36                            |
| 2                            | 21.1                       | 122.0                              | 3.79                            |
| 3                            | 17.4                       | 82.3                               | 3.12                            |
| 4                            | 14.9                       | 61.0                               | 2.68                            |
| 5                            | 13.4                       | 49.1                               | 2.41                            |
| 6                            | 12.3                       | 41.2                               | 2.20                            |
| 7                            | 11.3                       | 35.2                               | 2.04                            |
| 8                            | 10.6                       | 30.5                               | 1.90                            |
| 9                            | 10.0                       | 27.4                               | 1.80                            |
| 10                           | 9.5                        | 24.6                               | 1.70                            |

**Table 2** Closest radio-quiet quasar–galaxy pairs

| Coordinates  | Quasar  | Galaxy  | $m_G$  | $z_{QSO}$ | $r^*$ (arc min) |
|--------------|---------|---------|--------|-----------|-----------------|
| 1 0151 + 045 | PHL1226 | IC1746  | 15.1   | 0.404     | 1.8             |
| 2 0846 + 512 | W1      | Anon    | ~ 18   | †         | 0.2             |
| 3 1108 + 285 | QS      | NGC3561 | 14.7   | 2.192     | 0.7             |
| 4 1132 + 472 | BSO1    | Anon    | 15.1   | 1.13      | 1.7             |
| 5 1222 + 102 | Wdm6    | NGC4380 | 13.4   | (Cont.)   | 1.5             |
| 6 1233 + 125 | Wdm8    | NGC4550 | 12.5   | 0.728     | 0.7             |
| 7 1344 + 440 | BSO1    | NGC5296 | 15.0   | 0.963     | 0.9             |
| 8 1432 + 489 | QS      | NGC5682 | 15.1   | 1.94      | 1.7             |
| 9 2157 – 133 | BSO1    | IC1417  | ~ 15.0 | 0.71      | 1.3             |

\* $r \leq 3$  arc min, from Arp, ref. 11, except for QSO1108 + 285 (ref. 20).

†Blue variable star-like object.

(4) The choice of the galaxy may introduce two more observational selection effects in these small probabilities. Such probabilities are, in fact, the product of  $P_1 \times P_2$ ,  $P_1$  being the probability of finding a variable or blue stellar object in a particular location with respect to a galaxy<sup>19</sup>, and  $P_2$  the probability that this object is a quasar, equal to the number of pairs over the number of galaxies investigated. As the first step corresponding to  $P_1$  is omitted,  $P_2$  (a few per cent, according to ref. 11) is wrongly compared with the very small (but *a posteriori*) probability by chance ( $\sim 10^{-3}$ ) for such configurations, to prove the unlikelihood of the association; unfortunately the number of galaxies not investigated because of the absence of variable or blue stellar objects is inaccessible. A possible answer is that 'associated' quasars are associated only with galaxies having well-defined characteristics (morphological type, for example). Then the number of pairs involved here is much smaller and the existence of the excluded pairs can only be explained by chance, which proves then that close associations are not as rare as believed.

If a quasar lies within  $r$  arc min of a bright galaxy in the vicinity of which other fainter galaxies lie, (in a group for example, with a density  $\mu'$ ), then in  $P = 0.61 \mu' r^2$  of the cases, there will be another galaxy lying within  $r$  arc min of this galaxy and located nearer the quasar than the parent galaxy ( $P = 8.5\%$  if  $r = 10$  arc min and  $\mu' = 5(\text{deg})^{-2}$ ) and in  $P' = 1 - \exp(-\pi \mu' r^2)$  cases, another galaxy will lie within  $r$  arc min of the quasar ( $P' = 35\%$  with  $r$  and  $\mu'$  as above). In such cases the final angular separation will be

calculated from the fainter galaxy instead of the bright one around which the search was made (cases of quasars associated with companion galaxies<sup>11</sup>). Such situations, which are not very rare according to the indicative probabilities found here, can artificially increase the number of close associations and reduce the angular distances in the same way.

It is difficult to evaluate quantitatively the result of such observational selection effects, except by a careful systematic search for radio-quiet quasars around galaxies. But if we take these effects into account, as well as the estimated density of quasars and the expected distribution of angular distances, the list of very close quasar–galaxy pairs and the values of the probabilities are not as impressive as before. So far as we can conclude, the small angular distances found between some radio-quiet quasars and galaxies do not disagree with the expected values corresponding to the most likely values of the surface density of quasars. These spectacular associations correspond only to the toes of the distribution of the angular separations and the discovery of such quasars seems to be due only to their exceptional location.

I thank D. Wills for helpful discussions and Dr H. Arp for communicating data before publication. This work was supported by a grant from the French Foreign Office.

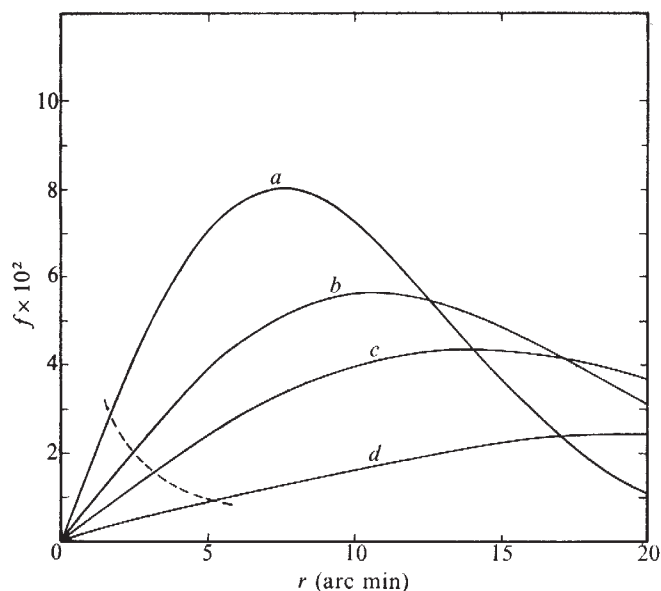
JEAN-LUC NIETO

Department of Astronomy,  
University of Texas,  
Austin, Texas 78712

Received 24 June; accepted 13 October 1977.

1. Arp, H. C., Baldwin, J. A. & Wampler, E. J. *Astrophys. J.* **198**, L3–L5 (1975).
2. Arp, H. C., Willis, A. G. & de Ruiter, H. R. *IAU Circ. No.* 2750 (1975).
3. Arp, H. C., Pratt, N. & Sulentic, J. W. *Astrophys. J.* **199**, 565–585 (1975).
4. Arp, H. C., Sulentic, J. W., Willis, A. G. & de Ruiter, H. R. *Astrophys. J.* **207**, L13–L16 (1976).
5. Arp, H. C. *Astrophys. J.* **210**, L59–L61 (1976).
6. Burbidge, G. R., O'Dell, S. L. & Strittmatter, P. A. *Astrophys. J.* **175**, 601–611 (1972).
7. Hazard, C. & Sanitt, N. *Astrophys. Lett.* **11**, 77–82 (1972).
8. Bahcall, J. N., McKee, C. F. & Bahcall, N. A. *Astrophys. Lett.* **10**, 147–152 (1972).
9. Browne, I. W. A. & McEwan, N. J. *Mon. Not. R. astr. Soc.* **162**, 21P–24P (1973).
10. Burbidge, E. M., Burbidge, G. R., Solomon, P. M. & Strittmatter, P. A. *Astrophys. J.* **170**, 233–240 (1971).
11. Arp, H. C. *IAU Colloq.* **37**; *C.N.R.S. Colloq.* **263** (1976).
12. Sandage, A. R. & Luyten, W. J. *Astrophys. J.* **155**, 913–918 (1969).
13. Bolton, J. G., Peterson, B. A., Wills, B. J. & Wills, D. *Astrophys. J.* **210**, L1–L3 (1976).
14. Walker, M. F., Pike, C. D. & Wills, B. J. & Wills, D. *Astrophys. J.* **210**, L1–L3 (1976).
15. Nieto, J.-L. thesis, Univ. Paris (1976).
16. Burbidge, G. R., Crowne, A. H. & Smith, H. F. *Astrophys. J. Suppl.* **33**, 113–188 (1977).
17. Bond, H. E., Kron, R. G. & Spinrad, H. *Astrophys. J.* **213**, 1–7 (1977).
18. Osmer, P. S. & Smith, M. G. *Astrophys. J.* **213**, 607–618 (1977).
19. Weedman, D. W. *Astrophys. Lett.* **9**, 49–51 (1971).
20. Stockton, A. *Astrophys. J.* **155**, L141–L142 (1969).

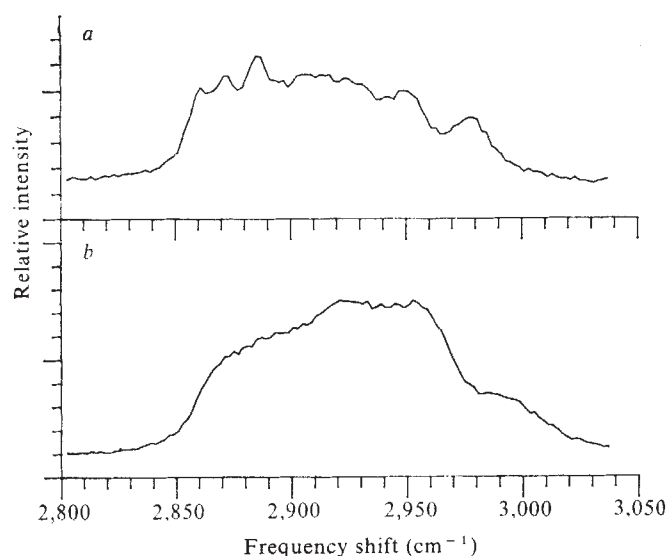
**Fig. 1** Distribution curves of the angular separations for different values of the density of quasars in  $(\text{deg})^{-2}$ . The dashed line indicates the locus of the lower limits of the 95% bilateral level range (see Table 1). a,  $\mu = 10$ ; b,  $\mu = 5$ ; c,  $\mu = 3$ ; d,  $\mu = 1$ .



## Pressure-induced changes in molecular conformation in liquid alkanes

LIQUID linear alkanes at room temperature and pressure exist in a mixture of molecular shapes. The *n*-heptane molecule, for example, has 13 possible distinct conformations with up to four gauche bonds distributed along its length. In the crystal, however, all the molecules assume the all-*trans*, straight chain shape. We have studied the conformations of chain molecules in the liquid state in various conditions using Raman scattering as a probe<sup>1</sup>. We have looked particularly at what happens to the shapes of linear alkane molecules subjected to high pressures in the liquid state. We speculated that as the pressure on a liquid of chain molecules was increased toward the freezing point that the chains might begin to straighten out. In fact, we found that an increase in pressure caused an increase in the number of gauche bonds, that is, pressure caused the molecules to become more globular.

Our interest in this investigation grew out of some recent discoveries concerning the conformation of linear polyethylene, a very long-chain molecule whose length can be as much as 1  $\mu\text{m}$  or more. Typically, polyethylene crystallised from the melt folds itself into straight chain segments tens of nanometres long, arranged side-by-side to form lamellae, with the segment axes perpendicular to the lamellar surface. The resulting material has relatively poor mechanical strength, presumably because



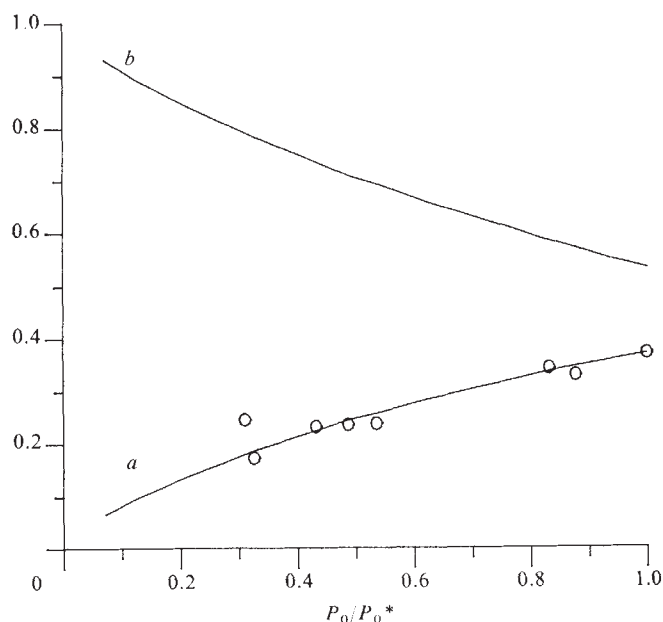
**Fig. 1** The CH stretching bands for heptane at 65 °C and pressures of *a*, 1 atm and *b*, 14.7 kbar. The peak at 2,880 cm<sup>-1</sup> disappears at high pressure as the molecules become more globular in shape.

it is held together mainly by Van der Waals forces between lamellae rather than by the carbon-carbon bonds of the polyethylene chain. On the other hand when polyethylene is crystallised at high pressure, ~ 5 kbar, the chains are extended to their full lengths in the solid<sup>2</sup> and, still more interesting, there is evidence that in the high pressure liquid before crystallisation the polyethylene chains exist in a 'nematic'-like state, an ordered liquid phase<sup>3</sup>. We wished to investigate this ordered phase and because the spectroscopy was more straightforward, decided to begin with short chain alkanes. Our results indicate that for the short chains the ordered liquid phase does not exist.

For heptane at room temperature and pressure the relative population of the all-*trans* conformer is about 15%<sup>4</sup>. The populations of other conformers can be calculated using the Boltzmann distribution, with each *gauche* bond requiring ~ 500 cal mol<sup>-1</sup> of energy. The greater statistical weight of the

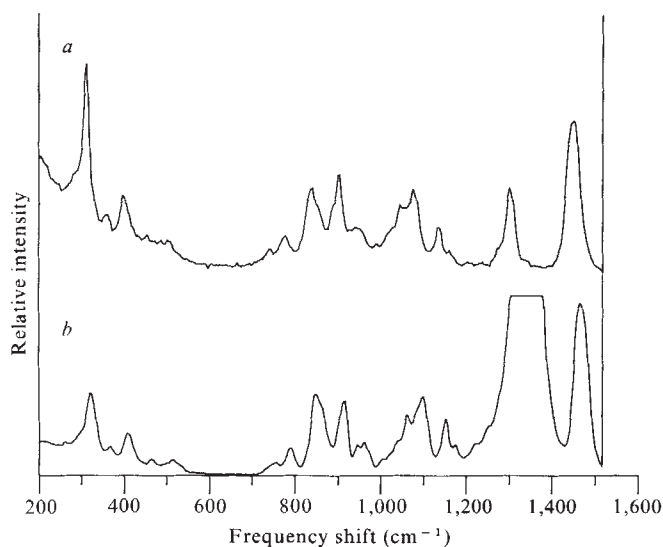
*gauche* conformers overcomes the energy barrier and for liquid alkanes longer than nonane the *trans* population is small and spectroscopic observation of the *trans* species is extremely difficult.

In the Raman spectrum of the alkanes there are two frequency domains that are especially conformation sensitive: the high frequency C-H stretching bands and the low frequency acoustic bands. The high frequency region for heptane at 65 °C is shown in Fig. 1. The upper spectrum was taken at room pressure and shows a fairly sharp peak near 2,880 cm<sup>-1</sup>, the intensity and



**Fig. 3** Relative populations of molecules with one *gauche* bond, *a*,  $P_1$ , and with more than one *gauche* bond, *b*,  $P_{2-4}$ . The abscissa is the all-*trans* population,  $P_0$ , relative to its room pressure value,  $P_0^*$ . Pressure is increasing to the left therefore. Solid curves are theoretical predictions; open circles are data points.

**Fig. 2** Spectra of heptane as in Fig. 1 but for the lower frequency range. The intensities of the acoustic bands in the region 300–600 cm<sup>-1</sup> drop as the pressure rises, indicating that the populations of molecules in the all-*trans* conformation and molecules with one *gauche* bond are decreasing in favour of more highly kinked conformers.



sharpness of which has been shown to be proportional to the crystallinity and packing density of neighbouring chain molecules<sup>5</sup>. In the lower curve taken in the liquid at 14.7 kbar pressure this band has disappeared, indicating a probable reduction in lateral chain order.

Figure 2 shows the spectral range 200–1,500 cm<sup>-1</sup> for the same two pressures for heptane. Most of this spectral range is only slightly affected by the change in pressure and the band at 1,450 cm<sup>-1</sup> was used as an intensity standard. (The pressures were generated in a diamond anvil cell. The strong feature appearing in the high pressure spectrum near 1,300 cm<sup>-1</sup> is the diamond Raman band.) The frequency range 300–600 cm<sup>-1</sup> contains the acoustic modes which are affected quite strongly by pressure. These bands have been the subject of much study<sup>6,7</sup>. Model calculations have been made, and some of the normal modes for different conformations have been identified with these bands<sup>8</sup>.

The sharp peak near 309 cm<sup>-1</sup> is the so called 'longitudinal acoustic mode' (LAM) which is due to a vibration of the all-*trans*, TTTT, conformers along their chain axes. Its frequency is inversely proportional to the chain length. The band at 507 cm<sup>-1</sup> results from a molecule with a single *gauche* bond, TGTT, the band at 363 cm<sup>-1</sup> is due to the GTTT conformer, and the band at 396 cm<sup>-1</sup> is caused by a superposition of bands from a single *gauche* bond conformation, GTTT, and a double *gauche* conformer, TTGG.

Note that in the room pressure spectrum the all-*trans* band at 309 cm<sup>-1</sup> is by far the strongest peak although it accounts for only 15% of the molecular population. It seems



that as the molecules become more globular in shape, vibrations along the chain length have less effect on the polarisability. The intensities of bands for 3- and 4-*gauche* bond conformers are too weak to be detected.

It is apparent from Fig. 2b that all the bands in the low-frequency region decrease as the pressure increases with the LAM dropping more rapidly than its neighbours. This shows that the populations  $P_0$ , of the all-*trans* molecules, and  $P_1$ , of the single-*gauche* bond molecules, are decreasing in favour of  $P_{2-4}$ , the population of the 2-, 3-, and 4-*gauche* bond conformers.

We have developed a theory, generalised from liquid-crystal theory<sup>9</sup>, which predicts this behaviour. Our model consists of a dense gas of hard rods interacting only through excluded volume effects and Van der Waals attraction. The rods have no orientational correlations and are free to rotate end over end as in the case of an isotropic liquid crystal. Different conformers of heptane correspond to different species of rods, each with its own aspect ratio (length to breadth ratio) and energy (depending upon the number of *gauche* bonds). If the aspect ratio of a rod decreases, the excluded volume swept out as it tumbles is reduced and the translational entropy increases. If we calculate and minimise the Gibbs free energy for this model we find that rods with lower aspect ratios become more favoured as the pressure and density increase. This means that the molecules become more globular as the pressure rises.

Figure 3 shows the predictions for the rise of  $P_{2-4}$  and the drop of  $P_2$  as functions of  $P_0/P_0^*$  ( $P_0^*$  is the room pressure value of  $P_0$ ). Pressure therefore is increasing to the left in this figure. The agreement between theory and experiment is gratifying.

We have also investigated the conformational behaviour of hexane, octane, and hexadecane. The first two have behaviour similar to that of heptane. We recall that there is evidence for an ordered phase in a polyethylene liquid at high pressure. Since such a phase does not seem to exist in the short chain liquids one might expect to find a 'crossover' chain length above which the ordered phase could be found. Hexadecane at room pressure is made up almost entirely of *gauche* bonded conformers. We found that as the pressure was increased no significant change in the conformational population could be detected spectroscopically up to the freezing point for temperatures up to 165 °C. We conclude that either hexadecane is shorter than the 'crossover' length or that the ordered phase is inaccessible in this temperature range. We are continuing our investigation with other alkanes and polyethylene.

P. E. SCHOEN  
R. G. PRIEST  
J. P. SHERIDAN  
J. M. SCHNUR

Naval Research Laboratory,  
Washington, DC 20375

Received 19 August; accepted 1 October 1977.

1. Schoen, P. E., Schnur, J. M. & Sheridan, J. P. *Appl. Spectrosc.* **31**, 337 (1977).
2. Wunderlich, B. & Arakawa, T. *J. Polymer Sci. A2*, 3697 (1964).
3. Monobe, K., Fujiwara, Y. & Tanaka, I. *Proc. Conf. High Pressure* 63 and 865 (Kawakita Kyoto, 1975).
4. Bartell, L. S. & Kohl, D. A. *J. chem. Phys.* **39**, 3097 (1963).
5. Gaber, B. P. & Peticolas, W. L. *Biochim. biophys. Acta* **465**, 260 (1977).
6. Snyder, R. G. & Schachtschneider, J. H. *Spectrochim. Acta* **19**, 85 (1963).
7. Snyder, R. G. *J. chem. Phys.* **47**, 1316 (1967).
8. Schaufele, R. F. *J. chem. Phys.* **49**, 4168 (1968).
9. Alben, R. *Molec. Cryst. Liq. Cryst.* **13**, 193 (1971).

## Particle size analysis for machinery health monitoring

MONITORING the health, or condition, of machinery is a universally accepted way of achieving economy of use and preventing catastrophic failure. Wear debris analysis is a powerful method of monitoring oil-lubricated machinery and many techniques are used, such as spectroscopic oil-analysis<sup>1</sup>, magnetic plug and filter inspection<sup>2</sup>, ferrography<sup>3</sup>, and particle size analysis. Although this

Table 1 Particle size analysis

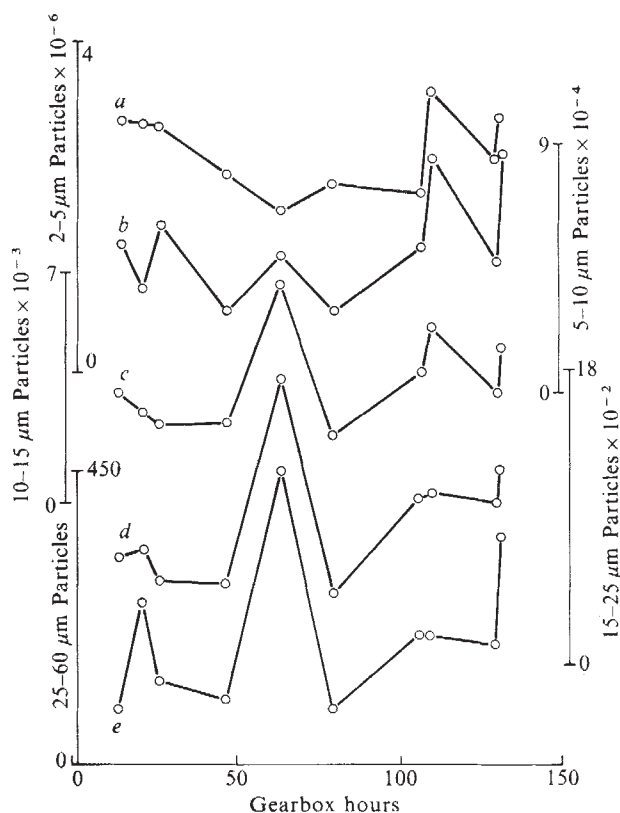
| Sample no. | Gearbox hours | <i>n</i> | <i>b</i> (μm) | <i>r</i> | Iron content (mg l <sup>-1</sup> ) |
|------------|---------------|----------|---------------|----------|------------------------------------|
| 1          | 15            | 3.183    | 0.907         | 0.955    | 7.0                                |
| 2          | 22            | 2.514    | 0.540         | 1.000    | 7.1                                |
| 3          | 26.75         | 3.068    | 0.842         | 1.000    | 7.9                                |
| 4          | 46.75         | 2.830    | 0.682         | 0.996    | 5.5                                |
| 5          | 63.25         | 2.382    | 0.725         | 0.994    | 5.5                                |
| 6          | 79.50         | 2.900    | 0.707         | 0.998    | 4.6                                |
| 7          | 106.75        | 2.773    | 0.843         | 0.997    | 5.0                                |
| 8          | 109.25        | 3.024    | 0.935         | 0.998    | 8.8                                |
| 9          | 129.75        | 2.744    | 0.745         | 0.997    | 6.9                                |
| 10         | 130.25        | 2.755    | 0.830         | 1.000    | 7.4                                |

These values of *n*, *b* and *r* were obtained with a lower size limit, *d'*, of 2 μm; this value corresponds to the limit of the sensor used.

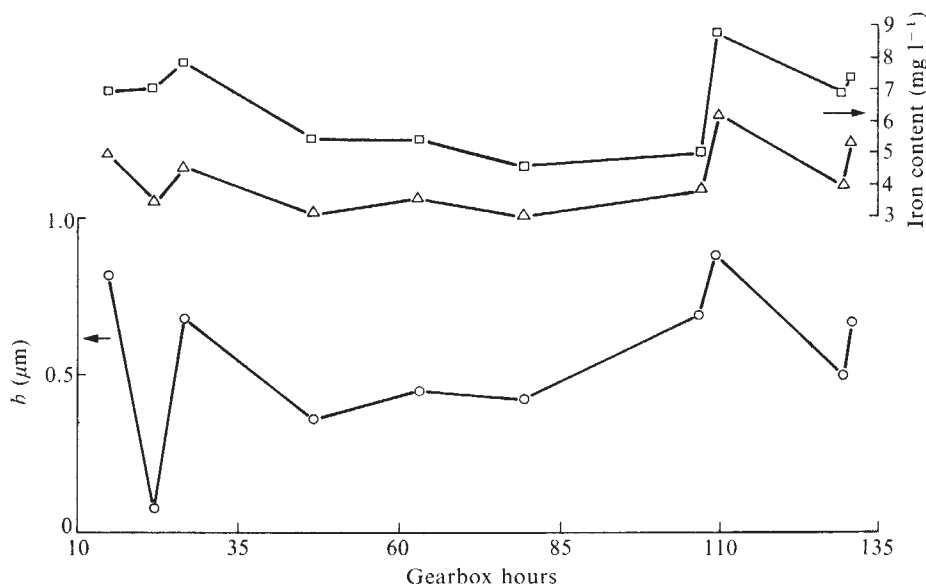
last technique is frequently used for monitoring hydraulic fluids, where contamination levels are low, it has not been accepted as a means of monitoring relatively dirty systems such as gas turbines where other techniques are more efficient and easier to interpret. Here we show how the results from particle size analysis can be reduced to a simple indicator of the severity of wear; this indicator is related to spectroscopic oil analysis and shows promise as a method of monitoring the new generation of ultra-clean gas turbines.

The cumulative distribution of wear particles as a function of particle size can be fitted to the distribution function,  $P(d) = \exp - [b/(d-d')]^n$ ;  $P(d)$  is the probability of a particle being smaller than *d*, *d'* is a lower particle size limit, *b* and *n* are the independent parameters which may be adjusted to fit the experimental results. This function is similar to that used by Rosin and Rammler<sup>4</sup> to describe the fineness of powdered coal. Taking logarithms twice

Fig. 1 Particle counts per 100 ml of oil for a, 2–5 μm; b, 5–10 μm; c, 10–15 μm; d, 15–25 μm; e, 25–60 μm. The apparatus used works on the principle of light blockage in which particles passing through a cell interrupt a light beam incident on a photo-diode. As the particle passes through the cell it tumbles and the maximum projected area is detected and the particle is sized as having the diameter of a sphere with the same cross-sectional area.



**Fig. 2** Values of  $b$  (○), measured iron content (□), and calculated iron content (△) as a function of gearbox hours. The calculated iron contents were determined from the integral  $T\rho\int_{20}^{250}\pi d^3/24p(d)dd$  where  $p(d) = dP(d)/dd$ ,  $T$  is the total number of particles per litre and  $\rho$  is the density of steel. The integral was evaluated using Simpson's rule.



yields  $\ln[\ln(1/P(d))] = -n\ln(d-d') + n\ln b$ . Thus a plot of  $\ln[\ln(1/P(d))]$  against  $-\ln(d-d')$  should be linear if the distribution function is applicable to the experimental results. From the form of  $P(d)$  we see that 36.7% of the particles will be smaller than  $b + d'$ . Since it is generally accepted that an increase in the relative number of large particles is an indication of an increasing severity of wear, it may be possible to use  $b$  as an indicator of the severity of wear.

Ten oil samples from a helicopter gearbox test-rig were used to test this theory. The particle sizes were analysed with a 'Hi-Ac' automatic particle size analyser; the iron contents were measured using an X-ray fluorescence spectrometer. The particle size analysis for the 10 samples is shown in Fig. 1. The values of  $n$ ,  $b$  and  $r$  (the regression coefficient) obtained by fitting the particle counts to the distribution function, and the measured iron contents are given in Table 1.

Figure 2 compares the values of  $b$ , the measured iron contents and the calculated iron contents. The close agreement between the trends in  $b$  and the measured iron content is striking and shows that  $b$  can be used as an indicator of wear. A comparison of Figs 1 and 2 illustrates how this analysis has simplified the somewhat complicated results of particle counting.

Figure 2 also shows the good agreement between the trends of the measured and calculated iron contents. Since wear particles are rarely spherical the expression for the iron content would be expected to be an overestimation. The good agreement actually observed is presumably due to the fact that particles smaller than  $2 \mu\text{m}$  are ignored by the particle counter.

I believe that this new interpretation of particle size analysis may well become an important technique for machinery health monitoring, especially as the trend towards improved filtration reduces the relative sensitivity of spectroscopic oil analysis.

Further details of this analysis as well as the relationship between particle size analysis and direct reading ferrography will be published elsewhere. I thank Mr P. Gadd of the Naval Aircraft Materials Laboratory for the particle size analysis.

G. POCKOCK

Admiralty Materials Laboratory,  
Holton Heath,  
Poole, Dorset, UK

Received 12 August; accepted 13 October 1977.

1. Bond, A. & Kittinger, D. C. *Proc. Ann. natn. Aerospace Electronics Conf.* Dayton 251-256 (1966).
2. Hunter, R. C. *Aircraft Engng* 4-14 (1975).
3. Bowen, R., Scott, D., Seifert, W. & Westcott, V. C. *Tribology Int.* 109-115 (1976).
4. Rosin, P. & Rammler, E. *J. Inst. Fuel* 7, 29-36 (1933).

## Intensification of the Gulf Stream after the winter of 1976-77

THE eastward volume transport of the Gulf Stream south of New England varies within a wide range. Within a triangle whose apex is at Bermuda and whose base extends from Cape Hatteras to Cape Sable, Nova Scotia, there have been 32 reliable oceanographic sections made across the Stream between 1932 and 1968. Geostrophic volume transports for these sections, computed relative to 2,000 m, are shown in Fig. 1. This 2,000 m surface is necessary if any seasonal change in transport is to be detected; only 10 of these sections reach to the ocean bottom ( $\sim 4,500$  m), and these were all made between April and August. These results tend to confirm an earlier suggestion<sup>1</sup> that the Gulf Stream is strongest in late winter and weakest in late autumn; this was attributed to the strengthening of the zonal winds in the winter. More recently, I have postulated a different mechanism for this annual variation in transport. The two most intense ocean currents, the Gulf Stream and the Kuroshio, are found on the western side of northern hemisphere oceans. These oceanic regions are visited throughout the winter by frequent outbreaks of frigid polar continental air. As a result of these outbreaks, very deep mixed layers are formed immediately south of these currents in late winter. The greatest thermocline depths are always found directly beneath the most deeply mixed layers, and it is primarily the variations in thermocline depth which account for the variations in computed transport shown in Fig. 1; the deeper the thermocline south of the Stream, the larger the transport. I suggest that the Gulf Stream receives a new charge of energy at the end of each winter due to this deepening of the thermocline. Here I discuss how the cold winter of 1976-77 influenced the Gulf Stream transport.

The temperature distribution in the western North Atlantic can be seen in the first more-or-less meridional section across the Sargasso Sea made by Thomson<sup>3</sup> in HMS Challenger. This section (Fig. 2) passed across the Gulf Stream in May; spring warming had started up at the sea surface above the deep, nearly mixed layer immediately south of the Gulf Stream. The Gulf Stream System is an asymmetric anticyclone. The narrow, intense eastward flow of the Gulf Stream itself takes place between

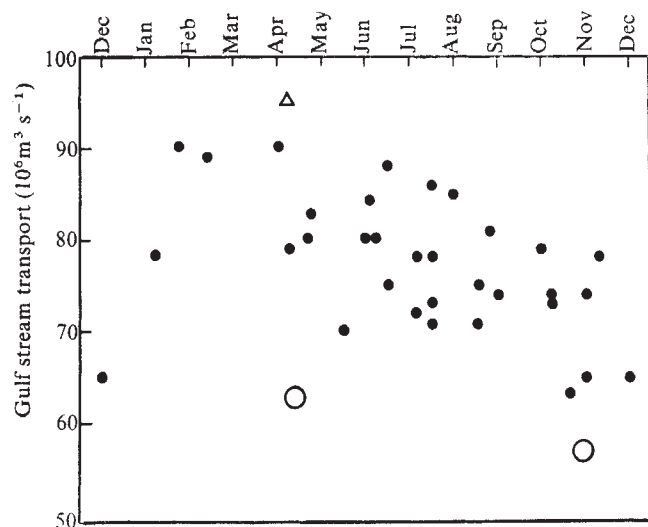


Fig. 1 Computed geostrophic volume transports for sections in the Bermuda - Nova Scotia - Cape Hatteras triangle: ●, 32 sections made between 1932 and 1968; ○, sections made before and after the winter 1974-75; △, between Researcher stations 6 and 33 in April 1977. All transports are relative to 2,000m and are plotted against month.

Challenger stations 42 and 44. The slower and much wider return flow toward the west takes place between stations 42 and 24, augmented by a further westward flow through the Caribbean Sea south of station 24. The maximum top-to-bottom transport of the anticyclone has been estimated to be about  $150 \times 10^6 \text{ m}^3 \text{ s}^{-1}$  (refs 4, 5) of which only about  $30 \times 10^6 \text{ m}^3 \text{ s}^{-1}$  pass through the Caribbean. Tropical surface water from the Caribbean can be seen as the 'warm core' at Challenger station 43. Superimposed on the anticyclone one can infer a shallow, meridional circulation in which excess  $18^\circ \text{C}$  water (ref. 6), formed at the end of the winter, flows south at the 300 m level and surface water is advected north to replace it. This process does not directly affect the Gulf Stream transport.

It can be seen (Fig. 2) that the thermocline is, in fact, deepest in the centre of the anticyclone beneath the thickest part of the lens of  $18^\circ \text{C}$  water. To the north, the thermocline rises abruptly across the Gulf Stream, and the lens of  $18^\circ \text{C}$  water disappears abruptly. To the south the thermocline rises gradually and the lens of  $18^\circ \text{C}$  attenuates gradually.

It is difficult to test hypotheses in the ocean, but an attempt was made during 1974-75 to observe the formation

of  $18^\circ \text{C}$  water and its possible effect on the anticyclone. An extensive oceanographic survey was made in October and November of 1974 and repeated in March 1975. No deep mixed layers were formed, and the  $18^\circ \text{C}$  water was not renewed by the atmosphere. Research ship cruises must be planned a year or two ahead of time, hence observers interested in short-scale climatic changes must simply take their chances. Precise calculation of the heat flux from ocean to atmosphere cannot be made at present since there is a three or four year lag before the necessary marine meteorological data can be obtained on magnetic tape, but the winter of 1974-75 was exceptionally mild over the northwestern Sargasso Sea.

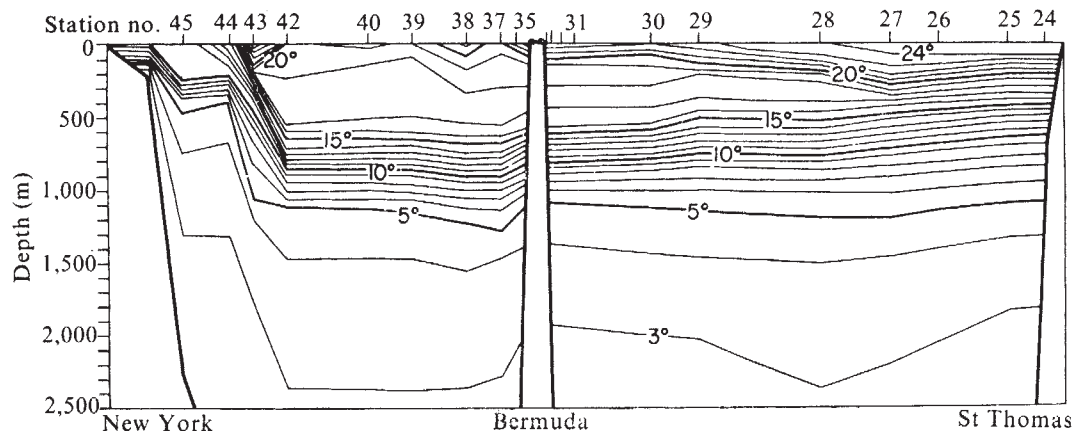
The winter of 1976-77, however, has been exceptionally cold in the eastern US. The surface atmospheric pressure map for mid-January to mid-February 1977 is drawn in Fig. 3. It is marked by a shift of the Icelandic low towards Newfoundland and a retreat of the Azores-Bermuda high towards the east. This was accompanied by a flow of continental polar air across the northwestern Sargasso Sea. This pattern was present much of the winter but broke up completely in March 1977.

I was invited aboard the Ocean Survey ship Researcher in April 1977 to investigate the effect of this remarkable winter on the  $18^\circ \text{C}$  water and its influence, if any, on the Gulf Stream transport.

Again, precise heat flux computations cannot be made at present, but the contrast between the mixed layer depths after the mild winter and the severe winter was vivid. Late winter temperature-depth and dissolved oxygen-depth curves for neighbouring oceanographic stations near the southern edge of the Gulf Stream are shown in Fig. 4. Knorr station 577 ( $34^\circ 31' \text{N}$ ,  $72^\circ 31' \text{W}$ ; 10 March 1975) has a mixed layer only 175 m deep. A faint inflexion can be seen in the  $18^\circ \text{C}$  water at 300 m. This is probably a relic of a previous winter when renewal of  $18^\circ \text{C}$  water by the atmosphere did take place. The concentration of dissolved oxygen at 300 m was 4.45 ml per l (81% of saturation), indicating that biological consumption of oxygen had been going on for some time. Below this inflexion the temperature gradient increases rapidly to  $1^\circ \text{C}$  per 50 m, the normal value for the thermocline in the Sargasso Sea. The depth of the  $10^\circ \text{C}$  isotherm, a commonly used index of thermocline depth, is 815 m.

The upper 200 m of Researcher station 33 ( $34^\circ 30' \text{N}$ ,  $71^\circ 18' \text{W}$ ; 12 April 1977) is composed of 'warm core' Gulf Stream water, but the rest of the top 600 m consists of newly formed  $18^\circ \text{C}$  water with oxygen values of about 5.1 ml per l (96% of saturation). The thermocline is much deeper; the  $10^\circ \text{C}$  isotherm lies at a depth of 1,020 m.

Fig. 2 Temperature section from St Thomas, Virgin Islands, to Bermuda to New York, made from HMS Challenger in spring, 1873.





deeper than at any of the earlier stations used for the computation of Gulf Stream transport (Fig. 1).

Computed volume transports for the two sections made before and after the mild winter of 1974–75 are also plotted in Fig. 1. While the late winter section ( $62 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ ) is slightly higher than the autumn section ( $57 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ ), these two sections give the two least volume transport figures ever computed, relative to the 2,000 m surface, in this area. Researcher station 33 has been paired with the nearest station north of the Gulf Stream (Researcher station 6;  $37^\circ 46' \text{N}$ ,  $72^\circ 00' \text{W}$ , 1 April 1977), and the computed volume transport between them,  $95 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ , has also been plotted in Fig. 1. It is the largest volume transport ever computed for the Gulf Stream.

With reservations, these data add credence to my proposed mechanism of oceanic anticyclonogenesis<sup>2</sup>. The

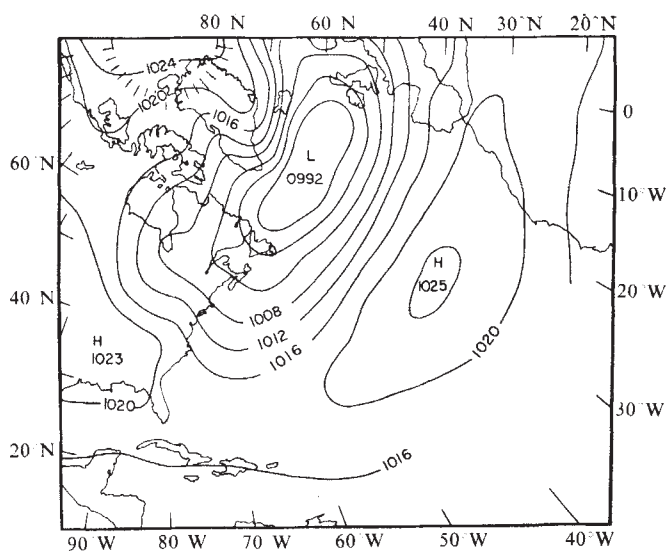


Fig. 3 Average surface atmospheric pressure for the period mid-January to mid-February 1977.

main reservations are these: Researcher was equipped with expendable bathythermographs that recorded the temperature down to 760 m. Thus it was possible to place station 33 in the exact centre of the region where the mixed layer was deepest and the thermocline was deepest. Knorr was similarly equipped during the two cruises which showed the lowest volume transports (Fig. 1) and these low figures are undoubtedly true. These sections were made, however, in a different longitude ( $64^\circ 30' \text{W}$ ) on the Nova Scotia side of the Hatteras–Bermuda–Nova Scotia triangle where the current (in the upper layers) may be weaker<sup>3</sup>. Also, the areas where the deepest mixed layers were found were not very wide. Such deep mixed layers could easily have been missed by earlier observers. It is further possible to assume that Researcher station 33 was in the centre of a small, local anticyclone independent of the Gulf Stream. Nevertheless, it seems that the Gulf Stream gained strength after the cold winter of 1976–77. I attribute this to deep winter overturn. It is hard to see how this intensification can have been due to wind-stress alone in view of the air-flow pattern indicated by surface atmospheric pressure (Fig. 3). The theoretical Gulf Stream transport corresponding to the mean wind-stress field is only  $23 \times 10^6 \text{ m}^3 \text{ s}^{-1}$  (see ref. 7). Thus the gross variability in baroclinic transport in the upper half of the Gulf Stream ( $57$ – $95 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ ) is nearly double the total mean theoretical wind-driven transport.

Masuzawa<sup>8</sup> has shown a meridional section across the Kuroshio system made in January 1967 in which deep

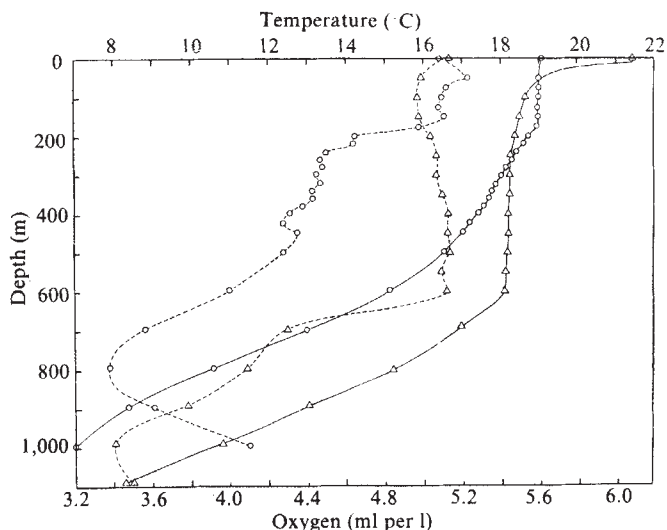


Fig. 4 Temperature–depth (solid lines) and oxygen–depth (dashed lines) curves for stations made south of the Gulf Stream in late winter 1975 (○) and early spring 1977 (△).

mixing may have contributed to the depression of the thermocline south of Japan. This section also shows the equatorward spreading of subtropical mode water, the northwest Pacific analogue of  $18^\circ \text{C}$  water. McCartney<sup>9</sup> has discovered that a similar process takes place on the equatorward side of the Antarctic Circumpolar Current. It seems that the greatest current systems in the world oceans are accompanied by deep mixing on their equatorward sides, and it is possible that they derive the greater part of their energy from this process.

This work was sponsored by the Office of Naval Research (Contract N00014-74C-0262, NR 083-004).

L. V. WORTHINGTON

Woods Hole Oceanographic Institution,  
Woods Hole, Massachusetts 02543

Received 6 June; accepted 1 October 1977.

1. Iselin, C.O'D. *Pap. Phys. Oceanogr. Met.*, 8, No. 1, 1–40 (1940).
2. Worthington, L. V. In *Studies in Physical Oceanography—a tribute to Georg Wüst on his 80th birthday* (ed. Gordon, A. L.) 1, 169–178 (Gordon and Breach, New York, 1972).
3. Thomson, C. W. in *The Atlantic*, 1, 1–424 (Macmillan, London, 1877).
4. Fuglister, F. C. *Prog. Oceanogr.*, 1, 265–373 (Pergamon, London, 1963).
5. Worthington, L. V. *The Johns Hopkins Oceanographic Studies* no. 6, 110 (The Johns Hopkins University Press, Baltimore, 1976).
6. Worthington, L. V. *Deep-Sea Res.* 5, 297–305 (1959).
7. Welander, P. in *The Atmosphere and the Sea in Motion* (ed. Bolin, B. B.) 95–101 (Rockefeller Institute Press and Oxford University Press, New York, 1959).
8. Masuzawa, J. in *Kuroshio: Its Physical Aspects* (eds. Stommel, H. & Yoshida, K. 95–127 (University of Tokyo Press, Tokyo, 1972).
9. McCartney, M. S. *Deep-Sea Res. Suppl.* 24 (in the press).

## Manganese-rich particulate matter in a coastal marine environment

PARTICULATE matter in the waters of the Gulf of St Lawrence, a major marginal sea, contains as much as 16,500 p.p.m. manganese. This is comparable with the levels of manganese found in deep-sea sediments<sup>1</sup>. Here we show that it is possibly a related phenomenon. Suspended particulate matter was collected by filtration on  $0.4\text{-}\mu\text{m}$  Nucleopore filters on a series of six stations along the axis of the Laurentian Trough. The samples were digested in a Teflon bomb with a mixture of aqua regia and hydrofluoric acid, and analysed by atomic absorption spectrophotometry<sup>2</sup>. Figure 1 shows a typical vertical profile of the concentration of suspended particulate matter and its content of manganese. The manganese content increases

rapidly with depth to a maximum value of 16,500 p.p.m. at 40–120 m above the bottom. The profiles are similar for the two cruises.

Resuspension of manganese-rich bottom sediment, and diffusion of dissolved manganese out of the sediment with subsequent precipitation in the water column, are two mechanisms which may produce manganese-rich particulate matter. Figure 2 shows that a manganese-rich surface layer, only a few millimetres thick, covers a sediment which otherwise is uniformly low in manganese. Higher concentrations of particulate matter near the bottom than at intermediate depths exist throughout the deep parts of the Gulf of St Lawrence<sup>3</sup>, indicating that resuspension of bottom sediments is an active mechanism<sup>4</sup>. But, the manganese content of the near-bottom particulate matter is higher than in the upper 1 mm of the sediment (by a factor of 4 to 5). Resuspension as the sole responsible mechanism requires, therefore, that the particulate matter retained in suspension—that is the slow settling components of the resuspended sediment—contain much more manganese than the bulk sediment. This possibility cannot be confirmed, however, because there are no data available on the manganese content of these sediments as function of settling rates.

The thinness of the manganese-rich sediment layer, and by inference the oxidised zone, indicates that manganese dissolved in the pore water would only have to diffuse a short distance to pass from the reduced zone to the water column. Subsequent precipitation would then result in manganese-rich particulate matter; this mechanism has been proposed by Evans *et al.*<sup>5</sup>. Graham *et al.*<sup>6</sup> have shown that diffusion of dissolved manganese out of sediments is fast enough to be a major factor in the manganese balance in Narragansett Bay.

Some of the manganese-rich particulate matter may reach the seaward-moving layer of the water-column by vertical mixing and advection, and ultimately find its way to the open ocean. This is supported by the data of Bowers and Yeats<sup>7</sup>, which show that whereas 93% of the input of particulate matter to the Gulf of St Lawrence sediments internally, only 47% of the manganese input is lost in this

Fig. 1 Vertical distribution of the concentration and manganese content of suspended particulate matter (SPM) on a deep station in the Laurentian Trough, Gulf of St Lawrence (48° 42.2' N, 68° 39.5' W). a, Manganese in May; b, manganese in September; c, SPM in May; d, SPM in September.

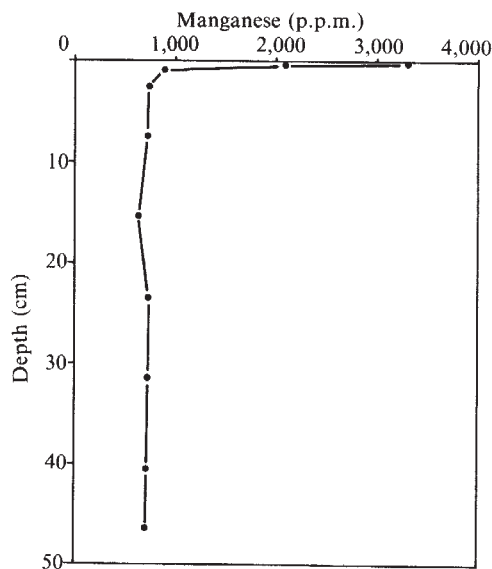
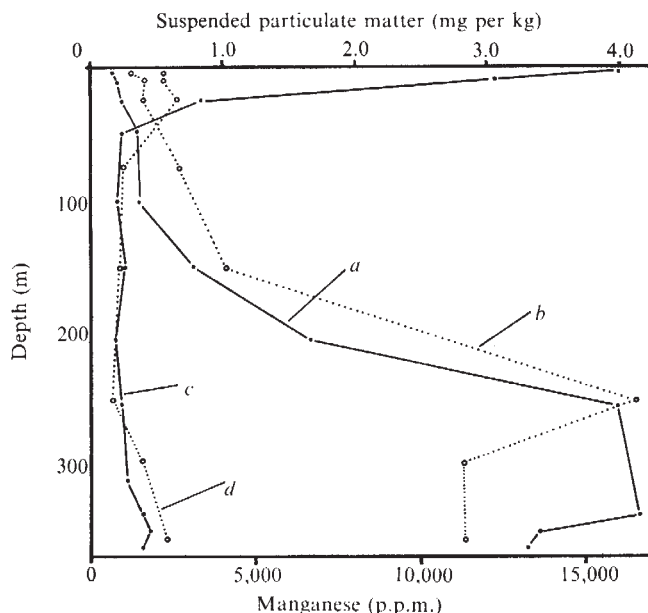


Fig. 2 Vertical distribution of manganese in a sediment core from the Laurentian Trough, Gulf of St Lawrence (48° 40.8' N, 68° 38.8' W). The core was taken with a box corer, subsampled immediately, and analysed by the method of Rantala and Loring<sup>17</sup>.

way. As the exported manganese is predominantly in particulate form, it seems that the exported particulate matter is enriched in manganese. Furthermore, the existence of individual particles, greatly enriched in manganese, has recently been demonstrated in suspended matter from the Atlantic Ocean<sup>8</sup>. Using scanning electron microscopy and an electron probe, it was shown that the particulate matter is composed of several phases of different manganese concentrations, so that the bulk composition is controlled by the relative abundance of these phases. Thus, manganese was enriched to about 10 times the normal level in clay particles that make up 10% of the total clay.

If production and export of manganese-rich particulate matter occurs generally in the coastal zone, it will have interesting implications for oceanic geochemistry. Manganese and other trace elements are not homogeneously distributed in deep-sea sediments. Thus, the highest concentrations are not found in clays from the abyssal plains adjacent to the continents, but from elevated areas remote from the continents. To explain this, Turekian<sup>9</sup> distinguished between two sediment transport mechanisms: bottom transport and deposition, and eupelagic deposition. The continentally remote hilltops contain material derived from the higher parts of the water column, that is the eupelagic components, and the abyssal plains adjacent to the continents contain the bottom transported components in addition. If very fine particles, rich in the critical trace elements, were transported by the surface layers of the oceans, then a uniform 'rain' of trace elements in particulate form would occur over the entire ocean. The bottom transported components would act as a diluent, causing the lower levels of trace elements observed in sediments adjacent to the continents.

The major problem with the theory is the origin of manganese-rich particles. Turekian<sup>9</sup> suggested that the fine particles could either be manganese oxide produced during the weathering of rocks, or small particles with high specific surface area which had absorbed the critical trace elements during stream transport. The difficulty with this is the implication that the removal of trace elements such as Ni, Co and Mn from seawater is of secondary importance to their removal in the river environment. The removal

of trace elements from seawater, and the maintenance of steady-state conditions, is thought to occur through scavenging by particulate matter settling through the water column<sup>10,11</sup>, and in particular by hydrous oxides of manganese in view of their high capacity for absorbing metal ions<sup>12</sup>. As Turekian<sup>9</sup> has pointed out, it is doubtful if particles transported by streams, be they manganese oxide or clay particles, are efficient scavengers of trace metals from seawater. As the concentrations of most trace elements are as high in streams as in seawater, trace elements would most likely have been absorbed to capacity on the stream particles. The higher sodium content of seawater would tend to inhibit further adsorption of trace elements from seawater, if not actually cause their release.

A marine, rather than river, origin of manganese-rich particles would remove this difficulty. These particles, with their content of freshly precipitated manganese oxide, would be the ideal scavenger, helping maintain the low concentrations of dissolved trace metals in seawater.

Another implication comes from massbalance calculations which involve manganese. Elderfield<sup>13</sup>, in a recent attempt to draw up a budget for oceanic manganese, concluded that the rate at which rivers supply dissolved manganese to the ocean is about an order of magnitude less than the rate at which non-lithogenous manganese accumulates in pelagic sediments. Hence, other significant sources of non-lithogenous manganese must exist.

One such source, if liberated by near-shore diagenetic processes, could be the non-lithogenous fraction in riverborne detritus, that is absorbed and complexed manganese and manganese oxide coatings. The riverborne detritus, of which some 95% never reaches the open ocean<sup>14</sup>, could then be considered a reservoir for manganese, available for remobilisation. The annual input of riverborne detritus to the marine environment is about  $2 \times 10^{16}$  g (ref. 15). If, from the total manganese carried by the detritus (estimated at an average of 4,300 p.p.m. (ref. 16)) as little as 50 p.p.m. were on average remobilised, that alone would amount to a flux of  $10^{12}$  g yr<sup>-1</sup>. This is an order of magnitude higher than the dissolved manganese supplied by the rivers, indicating the potential significance of this source.

BJORN SUNDBY

Section d'Océanographie,  
Université du Québec à Rimouski,  
Rimouski, Québec, Canada

Received 16 September; accepted 7 October 1977.

1. Goldberg, E. D. & Arrhenius, G. *Geochim. cosmochim. Acta* **13**, 153–212 (1958).
2. Rantala, R. R. & Loring, D. H. *At. Absorp. Newslett.* **16**, 51–52 (1977).
3. Sundby, B. *Can. J. Earth Sci.* **11**, 1517–1533 (1974).
4. Feely, R. A. *Mar. Chem.* **3**, 121–156 (1975).
5. Evans, D. W., Cutshall, N. H., Cross, F. A. & Wolfe, D. A., *Estuar. Coast. Mar. Sci.* **5**, 71–80 (1977).
6. Graham, W. F., Bender, M. L. & Klinkhammer, G. P. *Limnol. Oceanogr.* **21**, 665–673 (1976).
7. Bowers, J. M. & Yeats, P. A. *Nature* **268**, 595–598 (1977).
8. Chesselet, R., Spencer, D. W. & Biscaye, P. E. *Special Symp. Geochemistry and Ocean Mixing. Joint Oceanographic Assembly* (Edinburgh, 1976).
9. Turekian, K. K. in *Progress in Oceanography* (ed. Sears, M.) 227–244, (Pergamon, Oxford, 1967).
10. Goldberg, E. D. *J. Geol.* **62**, 249–265 (1954).
11. Turekian, K. K. *Geochim. cosmochim. Acta* **41**, 1139–1144 (1977).
12. Jenne, E. A. *Am. Chem. Soc., Adv. Chem. Ser.* **73**, 337–387 (1968).
13. Elderfield, H. *Mar. Chem.* **4**, 103–132 (1976).
14. Broecker, W. S. *Chemical Oceanography* 97–103 (Harcourt, Brace Jovanovich, New York, 1974).
15. Goldberg, E. D. *Comments Earth Sci. Geophys.* **1**, 117 (1971).
16. Aston, S. R. & Chester, R. in *Estuarine Chemistry* (eds Burton, J. D. & Liss, P. S.) 37–52 (Academic, London, 1976).
17. Rantala, R. R. & Loring, D. H. *At. Absorp. Newslett.* **14**, 117–119 (1975).

0.1 to 2.1 %<sup>2–6</sup>. Jacobs and Keeney placed river sediments mixed with inorganic mercury in the Wisconsin and Fox Rivers<sup>7</sup>. After 12 weeks exposure to the natural environment the concentration of methylmercury in the sediments was only 3%. This concentration of methylmercury was attained within 4 weeks after sediments were placed in the bottom of the rivers. This indicated that the processes of methylation and demethylation of mercury were in equilibrium in these environmental conditions within 4 weeks. But no experiments were conducted to show the degradation of methylmercury in sediments. After the methylation study by Jensen and Jernelöv<sup>8</sup>, many investigations into methylmercury production by microorganisms in bed sediments have been carried out. Little attention, however, has been paid to the ratio of methylmercury to the total mercury existing in the sediments. Spangler *et al.* observed degradation of methylmercury in mixed cultures from sediments<sup>9</sup>. The degradation reached 50% within 5 d, they also found methylmercury degradation in sediments. But detailed information about the sediments was not available and results fluctuated wildly. We do not know the total amount of methylmercury existing in all components of the systems, including fish, invertebrates and plants as well as various types of bed sediments. Here we use methylmercury equilibrium concentration levels in the Ottawa River sediments to estimate the amount of methylmercury in various types of bed sediment where detailed information of total mercury concentrations (total mercury estimation 31 kg) and types of sediment is available. Two ways (methylmercury production and degradation) of reaching equilibrium were observed during 50 d in identical environmental conditions.

Three types of fresh sediments from the study section of the Ottawa River were spiked with either radioactive mercuric chloride or methylmercuric chloride at a level of 1.00 p.p.m. The three sediments were: (A) main stream sediments, (B) north channel sediments, and (C) wood-chip sediments which were located in the north channel but just down-stream of the pulp and paper company. Original total mercury concentrations in the three types of sediments were determined by atomic absorption spectrometry as 17, 23 and 1,600 p.p.b., for type A, B and C sediment, respectively. Therefore, final concentrations of total mercury in type A, B and C sediments were 1,017, 1,023 and 2,600 p.p.b. More detailed information about these sediments is given elsewhere<sup>10</sup>.

Because no isotopic exchanges were observed between organic and inorganic mercury under natural environmental conditions and in these mercury concentration ranges<sup>11</sup>, it was possible to consider that increase or decrease of radioactive methylmercury in sediments was caused by methylation or demethylation not by simple isotopic exchanges. Organic contents of these sediments (A, B and C) were 0.35, 0.84 and 59.2%, respectively. The sediments were placed in the dark at 20° C and closed to prevent a loss of moisture as well as of mercury from the sediments.

Thin layer chromatography (TLC) and atomic absorption were used to determine the quantities of the various chemical forms of mercury, mainly mercuric and methylmercuric as well as the total mercury in sediments. A modification was made to extract various forms of mercury from sediments using 0.1% dithizone–benzene solution. Over 98% of the mercury from sediments was recovered using this modified method. *R<sub>f</sub>* values in the thin layer chromatography for mercuric and methylmercuric forms, respectively, were 0.36 and 0.66 using water–methylcellulose as the developing solvent. A portion of the mercuric and methylmercuric bands as well as other bands containing mercury were separated from the TLC sheet and were counted for radioactivity. All samples were analysed in duplicate.

Surprisingly, methylmercury in cleaner sediments (type A, main channel sediments (sands)) degraded more rapidly into inorganic forms, Fig. 1. Within 25 d, 80% of methylmercury was broken into inorganic forms in the sand sediments, whereas, only 14% of methylmercury was converted into

## Equilibrium concentrations of methylmercury in Ottawa River sediments

Most of the mercury (up to 97%) found in aquatic systems is associated with bed sediments<sup>1</sup>. Fish, invertebrates and plants contain only 0.02% of the total, though they contain a high proportion of organic mercury. Field surveys show that the proportion of organic mercury in the sediments range from



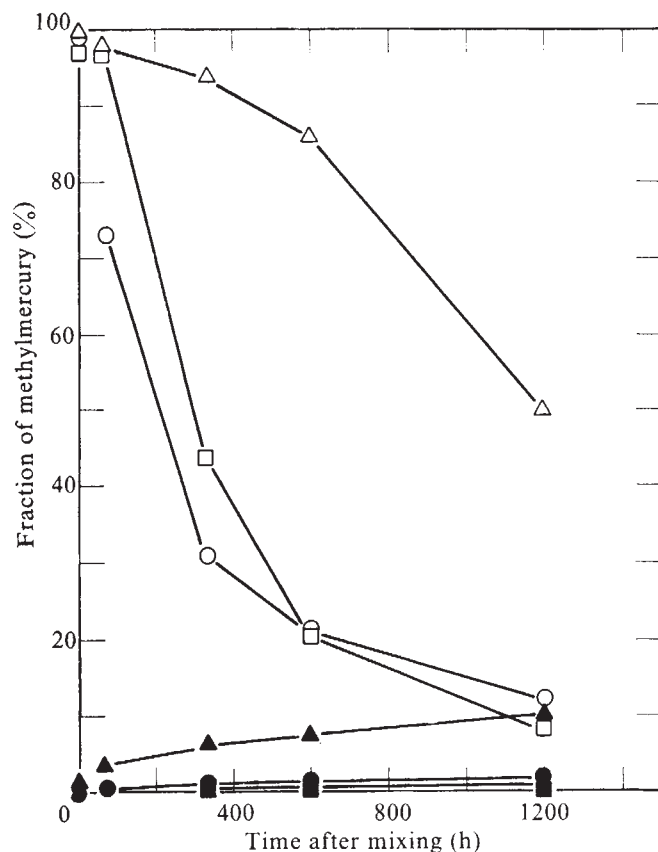


Fig. 1 Equilibrium concentration of methylmercury in Ottawa River sediments. Sediment A is shown by squares, B by circles and C by triangles, closed symbol, start from mercuric chloride only; open symbol, start from methylmercuric chloride only.

inorganic forms in the organic sediments (type C) during the same period. Sediments which were mixed with an inorganic form of mercury increased their contents of methylmercury with time. For highly organic contaminated sediments (type C, wood-chip sediments) the portion of methylmercury was 7.4% and for cleaner sand (type A, main channel sediments) the portion was only 0.1% after 25 d. The difference in final methylmercury concentration between sand (A) and wood-chip (C) sediments is probably a result of the different rates of formation and decomposition of methylmercury in these sediments. That is, the wood-chip sediment forms methylmercury more rapidly and also decomposes methylmercury more slowly than does the sand sediment. During the 50-d experiments, an average of 1.3% of mercury escaped from the system.

Methylmercury concentrations in each of the three types of sediments converged to an equilibrium concentration level from both directions starting with mercuric chloride and methylmercuric chloride, with an increase of time, Fig. 1. From the results, the equilibrium concentration levels were estimated for sediments A, B and C as 0.2, 2.3 and 12.1%, and the time to attain the equilibrium concentration was estimated as 3, 3, and 6 months, respectively.

Applying this methylmercury information to Ottawa River sediments, amounts of methylmercury were calculated as 1,200 g for the study section of the Ottawa River in 1972. In this calculation, the 4.88 km by 1.5 km section was divided into five regions depending on the type of bed sediments existing in the section of the river. The amounts of biomass in the study section were 5.3, 11.1, and 65.4 tons (wet) for fish, invertebrates, and plants, respectively. This biomass contained only 6.0 g of total mercury. In other words, bed sediments contained 200 times more methylmercury than those in biomass, even assuming all of the mercury in biomass was

methylmercury. Of course, the accuracy of this calculation of methylmercury in the sediments was not as high as that for the estimation of the total mercury in biomass in the study section. The calculation clearly showed, however, that methylmercury existing in the sediments was higher by at least two orders of magnitude, than those in the biomass. This further indicates that a detailed study must be directed toward mercury in bed sediments to understand and control, effectively, mercury pollution in aquatic systems.

We thank Mr J. Trepanier, Mrs E. Javorsky and Mr M. Y. Pinard for their contribution.

AKIRA KUDO  
HIROKATSU AKAGI  
D. C. MORTIMER  
DONALD R. MILLER

Division of Biological Sciences,  
National Research Council of Canada,  
Ottawa, Canada

Received 15 August; accepted 1 October 1977.

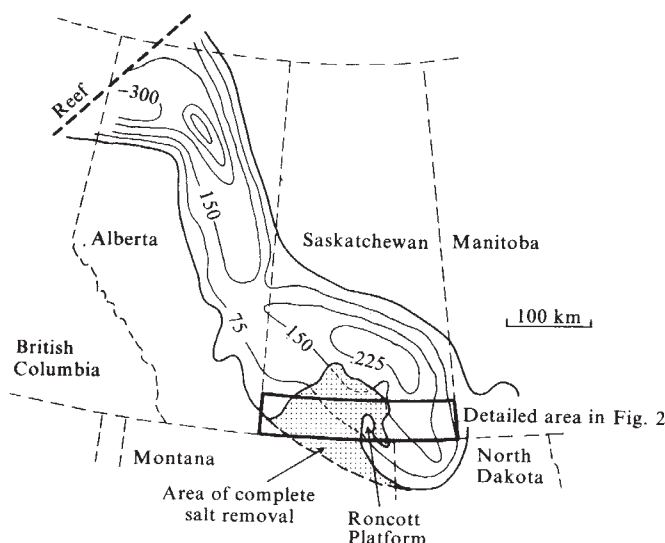
1. Kudo, A., Townsend, D. R. & Miller, D. R. *Prog. wat. Tech.* 9, 923 (1977).
2. Andren, A. W. & Harris, R. C. *Nature* 245, 256 (1973).
3. Batti, R., Magnaval, R. & Lanzola, E. *Chemosphere* 4, 13 (1975).
4. Langbottom, J., Pressman, R. & Lichtenburg, J. J. *Ass. Off. analyt. Chem.* 56, 1297 (1973).
5. Olson, B. H. & Cooper, R. C. *Nature* 252, 682 (1974).
6. Eganhouse, R. P. *Southern California Coastal Water Research Proj. Rep.* 83 (El Segundo, California, 1976).
7. Jacobs, J. W. & Keeney, D. R. *J. environ. Quality* 3, 121 (1974).
8. Jensen, S. & Jernelöv, A. *Nature* 223, 753 (1969).
9. Spangler, W. J., Spigarelli, J. L., Rose, J. M. & Miller, H. M. *Science* 180, 192 (1973).
10. Kudo, A., Townsend, D. R. & Miller, D. R. *J. Am. Soc. Civil Engng. Env. Div.* 103, EE4, 605 (1977).
11. Kudo, A., Akagi, H., Mortimer, D. C. & Miller, D. R. *Environ. Sci. Technol.* 11, 907 (1977).

## Origin of coal basins by salt solution

WILLISTON Basin of North America, one of the world's largest brown coal basins with diameter 550 km, is interpreted as an effect of regional solution of up to 150 m thickness of Devonian salt. The Late Cretaceous-Tertiary salt removal from the buried Elk Point Evaporite Basin (Middle Devonian) has resulted in sufficient collapse subsidence to be structurally reflected in the 2,000–3,000 m of overlying Palaeozoic and Mesozoic stratigraphic column. The Palaeocene coalfields are topset to Late Cretaceous delta lobes developed within salt solution troughs. This syndepositional mechanism for coal basin formation has not previously been recognised.

The most extensive coal basins were emplaced marginal to the Roncott Platform<sup>1-3</sup>, a major salt salient into the 2,500 km<sup>2</sup> of complete salt removal across southern Saskatchewan (Figs 1 and 2). Ground water circulation believed to have caused the salt solution is attributed to flow under hydraulic head from the nearby Rocky Mountains which were emerging to the south-west. Initially, it was directed along vertical planes associated with rejuvenated Precambrian lineaments, which are related to a possible plate boundary<sup>4</sup> that trends across western South Dakota and eastern Montana and traceable into south-central Saskatchewan as the salt solution axes flanking the Roncott Platform.

The Palaeocene coal measures are coal-bearing bodies that are topset to the Cretaceous delta lobes. Wood Mountain-Willow Bunch complex is the largest<sup>5</sup> of the coalfields (Fig. 2). Its several depositional troughs are along the western margin of the Roncott Platform (Figs 2 and 3). The thicker parts of seams overlie anomalous negative troughs on the Late Cretaceous structural surface, which themselves partially represent syndepositional clastic thickening trends. The coal seams of the Wood Mountain and Willow Bunch coal basins have their thicker parts superimposed. These are in turn vertically above the Late Cretaceous delta lobes' axial lengths. Without the underlying solution subsidence any thick peat accumulation associated with temporary mobile axes would not be preserved



**Fig. 1** Isopachous map of the Prairie Evaporite Formation (Devonian) salt and potash beds in the Elk Point Evaporite Basin across western Canada. Contour interval is 75 m. The salt beds to the south-east were leached, resulting in evaporite-solution brecciation of superjacent strata, especially during the Late Cretaceous to Lower Tertiary. The resulting extensive collapse subsidence over this zone was reflected up the 2,000–3,000 m stratigraphic column during the emplacement of Williston Basin's deltaic topset coal basins, whose preservation was assured within the subsidence troughs. The best developed coal basins were emplaced marginal to the Roncott Platform, a salt salient that extends northwards from northeastern Montana into the leached zone (Fig. 2).

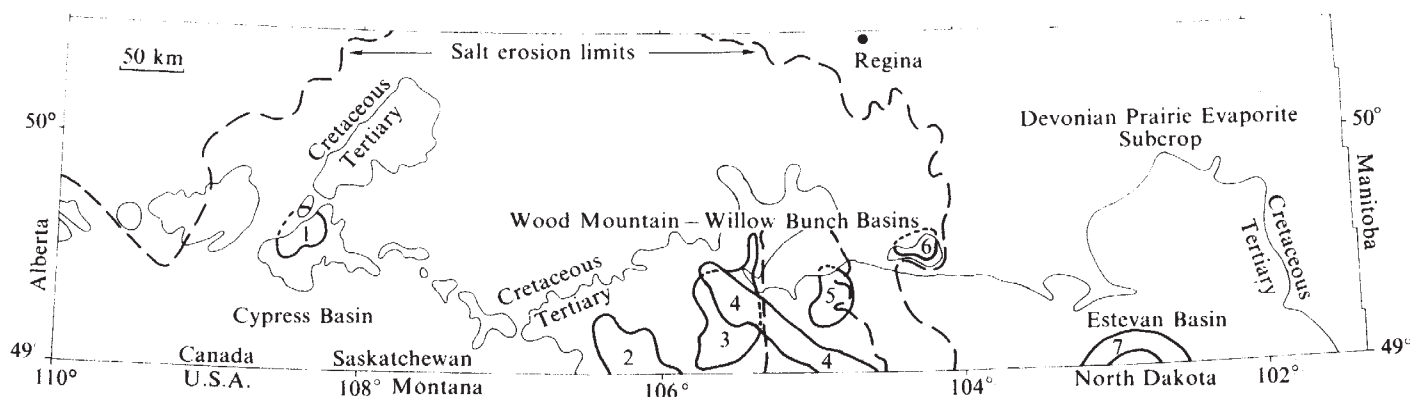
in the record. This phenomenon of coal basin genesis is well illustrated by deposition within the Coronach Trough, marginal to the western flank of the Roncott Platform (Fig. 2). There are six major seam groups in the Ravenscrag Formation of the Coronach Trough. The thickest seam of the 80–120 m coal measures in the trough, as well as in the formation, is the Hart seam. Its thickened trend, the 3–10 m thickness isopach interval, is a north–south subcrop belt, 8–24 km wide, that parallels the 65 km length of the Roncott Platform salt subcrop erosional edge north of the International Boundary (basin 3, Fig. 2). The Hart seam thins eastward in the subcrop along a broad front paralleling this salt erosional edge such that the

thinned depositional margin of the seam basin and the Roncott Platform western erosional edge coincide in the vertical succession along most of its length. Where the seam encroaches on to the salt platform subcrop zone, it is less than 1-m thick. All seams in the coal measures that are of sufficient thickness to be correlative between boreholes in the subcrop have their axes of maximum thickness superimposed, paralleling the Roncott Platform erosional margin and vertically along the maximum thickness trend of the Hart seam. They similarly thin eastward along a broad front as they encroach the salt substrate. The uppermost seam group of the Willow Bunch Basin, the 4–7 m thick seams of the Willow Bunch group (basin 4, Fig. 2) are exceptional in that they transect the Roncott Platform; their deposition axes are oblique to the Coronach Trough subsidence. The isopachous trends of Willow Bunch seams demonstrate the markedly diminished influence of salt solution subsidence emplacement and the assertion of the regional cratonic southeastern subsidence in the syntectonic sedimentological framework.

The salt solution subsidences of the Killdeer–Rockglen and Coronach Troughs (Fig. 3) were contemporaneous with the final recession of the Bearpaw sea. They favoured the fixation of vertically accreting Late Cretaceous deltaic lobes within a Catskill-like<sup>6</sup> muddy shoreline that transgressed over the sea. The main delta body, approximately 100 km across, extends south of the International Boundary beyond the study area (Fig. 3), and is comparable in size to the largest of deltas observed in the sedimentary record world-wide. The delta lobe accretion along subsidence axes inhibited lateral progradation and developed the Palaeocene coal measures as topset beds: the margins of the largest seam basins coincide with the subcrop extent of the late Cretaceous lobe. This sequence is repeated with modification eastward into the Hummingbird Trough along the eastern margin of the Roncott Platform (Fig. 3). Sufficient but incomplete removal of evaporite beds contemporaneous with the Late Cretaceous continental transgression across the eastern Platform developed a series of coalesced 5–10-km diameter lobes (Fig. 3) that encroached the western limb of the Hummingbird Trough. The relatively small Bengough coalfield (basin 5, Fig. 2) in the overlying Palaeocene thus is situated on this semi-stable Roncott salt platform area, but its thick seams do not laterally extend into the axial area of the Hummingbird Trough. This may be due to unfavourable subsidence for accumulation and preservation of peat sections.

The few major intracratonic deltas studied in the ancient sedimentary record, for example, the Upper Cretaceous delta of

**Fig. 2** The late Tertiary and Pleistocene drift blankets the unconformably Late Cretaceous to Early Tertiary bedrock surface on the northern limb of the Williston Basin, southern Saskatchewan. Most of the Palaeocene coal basins of this borehole exploration area are emplaced within salt solution subsidence zones: Cypress (1) Wood Mountain (2) and portions of the Willow Bunch (3–6) Basins. The lower coal measures of the Coronach Trough (3) are emplaced to offset to the paralleling western margin of the Roncott Platform's rigid salt substrate, whereas the Bengough coalfield (5) is developed with the partial solution subsidence of the eastern half of the salt structure. Minor coal basins (6) are emplaced within embayments into the salt substrate along the eastern Hummingbird Trough (Fig. 3) contact with the regional salt substrate. The stratigraphically high Willow Bunch seam group basin (4) is the least affected by the solution subsidence, and transects the subjacent solution axis. Likewise, the equivalently young Estevan Basin (7) coal measures develop with the assertion of cratonic subsidence as the dominate syndepositional genesis of coal basins in place of solution subsidence.



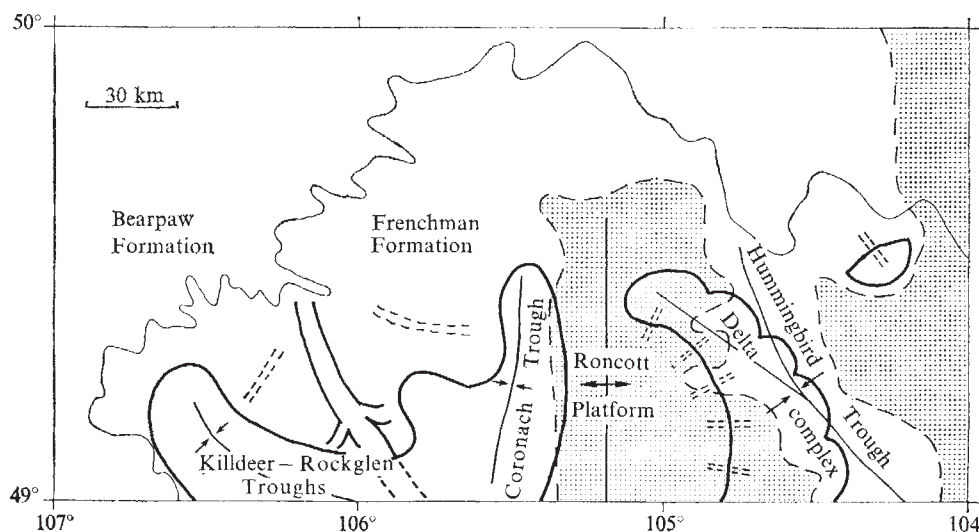


Fig. 3 Emplacement of the Late Cretaceous delta lobes along salt solution subsidence axes within the Frenchman formation. These lobes persisted in the vertical succession as topset Palaeocene coal basins. Extensive progradation of the main lobe along the Coronach axis was inhibited by the structural barrier of rigid salt substrate (stippled) of the western Roncott margin.

the central Rocky Mountains<sup>7-8</sup>, have been created by crustal downwarping reinforced by sediment loading. By contrast, solution subsidence as the cause of the Late Cretaceous to Lower Tertiary Wood Mountain-Willow Bunch delta complex suggests that loading is of minor importance. The crustal movement along the postulated geosuture was primarily lateral with minimal vertical component. Such movement does not preclude the development of vertical fracture zones into the Phanerozoic, including the Devonian salt beds to promote channelisation of sufficient fluids for regional dissolution.

The initiation of thick seams in the coal measures on the northern limb of the Williston Basin is transgressive southeastward towards the depocentre in northwestern North Dakota (Fig. 4). The thick seams of the Cypress Basin are at the base of the Tertiary (Ravenscrag Formation): the Ferris and superjacent Anxiety Butte seams, each developed to between 4 and 8 m thick. This interval is approximately isochronous with the relatively thinned Landscape seams, tens of metres higher in the stratigraphic section of the Coronach Trough, 200 km to the east. The Willow Bunch Basin subsidence did not diminish, to a rate permitting thick peat accumulation until the onset of the Hart seam 100 m higher in the section (Fig. 4). Similarly,

these conditions were not developed in the Estevan Basin, another 200 km to the east, until an additional 100 m of section was deposited above thinned and discontinuous equivalents to the Hart seam. The Boundary seam, the lowest of the thick seams in the Estevan Basin coal measures is isochronous on the basis of palynology with the main thick seam of the stratigraphically high Willow Bunch group across the Willow Bunch Basin (Fig. 4).

The onset of subsidence rates favourable for accumulation and preservation of thick peat sections is interpreted as a linear transgression up section as distal limb sections detach from the relatively rapid but diminishing subsidence of the cratonic basin. This is contemporaneous with the transition up section from subsidence dominated by salt solution to sufficiently diminished cratonic subsidence favourable for accumulation of thick peat sections in stagnant basins.

This work was supported by the Saskatchewan Department of Mineral Resources and the Geological Survey of Canada, 1971-77, and is based on a regional exploration programme involving the drilling of several thousand boreholes. I thank J. A. Irvine of the Institute of Sedimentary and Petroleum Geology (Calgary), and S. H. Whitaker of the Saskatchewan Research

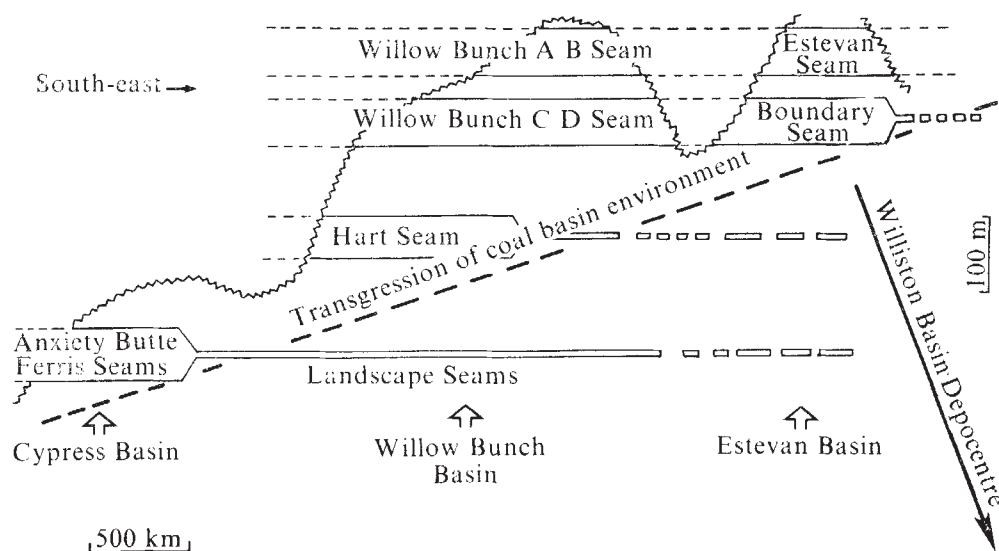


Fig. 4 Representation of the up section accumulation of the thick coal seams as a linear transgression of detached distal limb sections towards the craton depocentre.



Council (Saskatoon) for their helpful discussions of the borehole stratigraphy; P. Gulio of the Saskatchewan Geological Survey (Regina), P. F. Friend and A. G. Smith of the Department of Geology, University of Cambridge, for helpful comments on early versions of the text and A. Sweet of the Geological Survey of Canada (Calgary) for palynological correlations of various sections.

PAUL L. BROUGHTON

Department of Geology,  
University of Cambridge,  
Cambridge, UK

Received 12 August; accepted 3 October 1977.

- Holter, M. *Sask. Dep. Min. Res. Rep.* **123**, 113 (1969).
- Gorrell, H. & Alderman, G. *Geol. Soc. Am. Spec. Paper* **88**, 291–317 (1968).
- Smith, D. & Pullen, J. *Bull. Can. petrol. Geol.* **15**, 468–482 (1967).
- Camfield, P. & Gough, D. *Can. J. Earth Sci.* **14**, 1229–1238 (1977).
- Broughton, P., Irvine, J. & Whitaker, S. *Sask. Geol. Soc. Spec. Publ.* **2**, 81–94 (1975).
- Walker, R. & Harms, J. *J. Geol.* **79**, 381–399 (1971).
- Audley-Charles, M., Curran, J. & Evans, G. *Geology* **5**, 341–344 (1977).
- Weimar, R. *Soc. Econ. Paleont. Mineral. Spec. Publ.* **15**, 270–292 (1970).

## Primitive lead in an Australian Zn–Pb–Ba deposit

RADIOACTIVE decay of  $^{238}\text{U}$ ,  $^{235}\text{U}$  and  $^{232}\text{Th}$  yields  $^{206}\text{Pb}$ ,  $^{207}\text{Pb}$ , and  $^{208}\text{Pb}$ , respectively, as daughter products. This radiogenic lead is continually added to common lead in the source environments of ore deposits and the lead isotopic ratios of the resulting mixture are consequently time dependent. Because the relative proportions of these lead isotopes are fixed when the lead, common and radiogenic together, is separated from the parental uranium and thorium and incorporated into a lead deposit, determination of the isotope ratios provides evidence regarding the age of the lead ore and thus indirectly of the rock in which the ore is contained. Generally speaking, the older the ore the less radiogenic will be its lead. We report here analyses of extremely primitive lead from a small lead-bearing deposit in Western Australia.

The Big Stubby Zn–Pb–Ba deposits, located 5 km south of Marble Bar in the Pilbara Goldfield of Western Australia, have recently been described by Reynolds *et al.*<sup>4</sup> and their similarities to the Kuroko deposits have been discussed by Brook<sup>1</sup>. The conformable sulphide–sulphate deposits occur in an Archean volcanic sequence consisting of rhyolite breccias, rhyolite, and intermediate volcanics. Sphalerite and galena, dispersed within a barite gangue, are associated with a chert layer overlying a number of small rhyolite domes. The entire volcanic belt is intruded by granitic rocks dated at  $3.125 \pm 366$  Myr (ref. 3).

Five individual sulphide lenses are known at Big Stubby but only one, the 'A' Lens, has been systematically drilled where the best intersection is 4.27 m of 0.34% Cu, 4.5% Pb, 25.0% Zn, 1.071 g Ag per tonne and 15% Ba. Four intersections have been made in the 'A' Lens over a length of 500 m and the average grade is 0.26% Cu, 4.6% Pb, 16.5% Zn, 533 g Ag per tonne and 17% Ba.

It should be noted that the largest known barite deposits in Western Australia occur only 40 km WNW of Big Stubby and

**Table 2** Average isotopic compositions of lead from Old Star, Letaba and Big Stubby

| Deposit    | $^{206}\text{Pb}/^{204}\text{Pb}$ | $^{207}\text{Pb}/^{204}\text{Pb}$ | $^{208}\text{Pb}/^{204}\text{Pb}$ | Ref.      |
|------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------|
| Old Star   | 12.214                            | 13.851                            | 32.16                             | 5         |
| Letaba     | 12.679                            | 14.089                            | 32.51                             | 5         |
| Big Stubby | 11.95                             | 13.72                             | 31.94                             | This work |

have also been described as conformable deposits associated with sedimentary chert layers in an Archean volcanic sequence<sup>2</sup>.

Lead isotope analysis of three galena samples from Big Stubby are shown in Table 1. A significant feature of the Big Stubby lead is that it is the least radiogenic 'ore lead' reported to date, although feldspar from the Amitsoq gneiss of West Greenland contains lead which is even less radiogenic<sup>7</sup>. It is considerably more primitive than lead in galena from carbonate veins in the Old Star gold mine, South Africa, previously considered to contain the least radiogenic lead<sup>5</sup>. The same authors report that lead from Letaba, a subeconomic strata-bound massive pyritic deposit, although more radiogenic than that in the Old Star vein, is the least radiogenic found in a conformable deposit. It now seems that Big Stubby contains lead more primitive than either of these. Furthermore, an average of the Big Stubby data yields a model lead age of about 3,500 Myr according to the Stacey–Kramers model<sup>6</sup>.

For comparison, average isotopic compositions of lead from Old Star, Letaba and Big Stubby are presented in Table 2. Further lead and sulphur isotopic analyses of samples from this ancient segment of the earth's lithosphere are underway.

This article is published with the kind permission of Messrs S. H. and M. T. Stubbs of Marble Bar, W.A. and Aquitaine Australia Minerals Pty Ltd.

D. F. SANGSTER

Geological Survey of Canada,  
Ottawa, Canada

W. A. BROOK

Aquitaine Australian Minerals Pty Ltd,  
Western Australia 6000,  
Australia

Received 17 August; accepted 7 October 1977.

- Brook, W. A. *45th Annu. Congress, Perth Western Australia, August 13–17, 1973. Section 3, Geology, Symposium 8, Paper 8.8* (Abstr.) (1973).
- Hickman, A. H. in *Geological Survey Western Australia, Annual Report 1972* 57–60 (1973).
- de Laeter, J. R. & Blockley, J. G. *J. geol. Soc. Aust.* **19**, 363 (1972).
- Reynolds, D. G., Brook, W. A., Marshall, A. E. & Allchurch, P. D. *Aust. Inst. Min. Met. Mono Series No. 57* 185–195 (1975).
- Saager, R. & Koppel, V. *Econ. Geol.* **71**, 44–57 (1976).
- Stacey, J. S. & Kramers, J. D. *Earth planet. Sci. Lett.* **26**, 44–57 (1975).
- Baadsgaard, H., St. J. Lambert, R. & Krupicka, J. *Geochim. cosmochim. Acta* **40**, 513–527 (1976).
- Stacey, J. S., Delevaux, M. H. & Ulrych, T. S. *Earth planet. Sci. Lett.* **6**, 15–25 (1969).

## East Anatolian earthquake of 24 November 1976

A DESTRUCTIVE earthquake occurred north of Van in East Anatolia on 24 November 1976 at 12:22:18.3 GMT (USGS epicentre: 39.10°N and 44.02°E; average magnitude:  $M_s = 7.3$ ). The earthquake did extensive damage, destroyed more than 80% of the dwellings in a 2,000 km<sup>2</sup> area and caused more than 4,000 deaths. In this area of complicated geology<sup>1–3</sup>, this was the only known large earthquake in the last century according to seismicity catalogues<sup>4–5</sup> and the recollection of villagers. The fault trace and the displacement were clearly visible and could be mapped for the total length of the fault. Furthermore, any precursory phenomena that might have preceded the earthquake as perceived by the local residents along the whole fault could be investigated. We summarise here the results of extensive field investigations carried out during the first 10 days immediately following the earthquake, and again in July 1977.

The location of the earthquake relative to major faults and earthquake epicentres is shown in Fig. 1. The earthquake is to the

**Table 1** Lead isotope analysis of galena samples from Big Stubby

| Sample no. | $^{206}\text{Pb}/^{204}\text{Pb}$ | $^{207}\text{Pb}/^{204}\text{Pb}$ | $^{208}\text{Pb}/^{204}\text{Pb}$ |
|------------|-----------------------------------|-----------------------------------|-----------------------------------|
| Sp-3523    | $11.95 \pm 0.09$                  | $13.74 \pm 0.10$                  | $31.96 \pm 0.24$                  |
| SP-3525    | $11.97 \pm 0.09$                  | $13.74 \pm 0.10$                  | $31.93 \pm 0.24$                  |
| SP-3527    | $11.93 \pm 0.09$                  | $13.69 \pm 0.10$                  | $31.92 \pm 0.24$                  |

Analyses were performed by Teledyne Isotopes, Inc., under the direction of L. Casabona. The data are corrected to absolute values on the basis of measured and absolute values for standard lead SRM 981<sup>8</sup>. The approximate 95% confidence level, based on the analysis of 20 duplicate samples, is 0.76%.

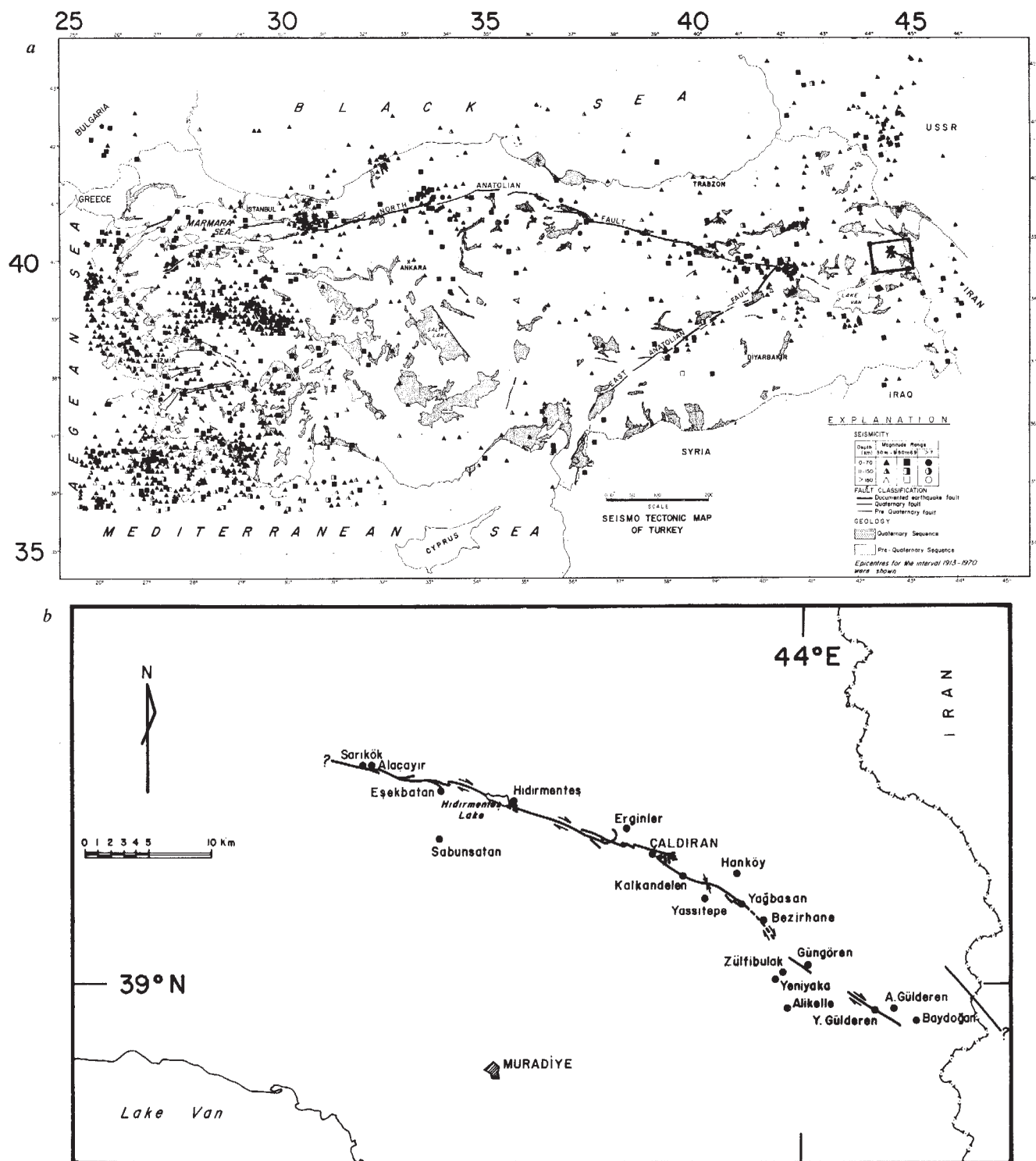


Fig. 1 a, Location of the earthquake within the seismotectonic setting of Turkey (inside the box), and b, detailed map of the observed fault trace.

east of the junction of the North Anatolian and East Anatolian Faults. The observed fault trace is to the north of a possible southeastern extension of the N. Anatolian fault trace through Lake Van. The connection of this earthquake fault with those in Iran is not clear, although there are a number of right-handed, strike-slip faults in Iran (branches of the North Tabriz fault) extending toward the Turkish border<sup>6</sup>. There are young volcanoes in the area and Tendurek, only 20 km north of the epicentre, has been active in the quaternary period. A more detailed description of the geology and tectonics of the region is given by Arpat *et al.*<sup>7</sup>.

The fault trace was clearly visible for about 50 km (Figs 1b, 2). The motion was almost purely right-lateral, strike-slip. The surface trace has a strike of about N 70°W in the central and western side and it bends somewhat east of Caldıran and continues along an azimuth of about 135°. The surface break is more prominent in the central and western sections, and has some *en-echelon* jumps in the east. The fault disappears abruptly under mountains in the west. At the eastern end the disappearance is gradual. The fault trace splinters and disappears toward the Iranian border. The observed horizontal displacements are more than 3 m (330–350 cm) in the west, about 250 cm in the central

regions and less to the east. The dip is nearly  $90^\circ$ . There are vertical displacements with inconsistent direction, generally about 50 cm, observed locally in several areas. The fault plane solution gives a right-handed, strike-slip source mechanism. The photographs in Fig. 2 show the fault trace and clearly document the right-handed motion. From the distribution of displacements, it seems that faulting may have started close to the west end and propagated to the east. The reported duration of severe shaking in Van was about 17 s, consistent with a fault length of about 50–60 km.

Seismic moment and stress drop can be calculated from the field observations. Taking the fault length  $L = 55$  km, assuming fault width  $W = 1/3 L$  and the average observed displacement  $D = 250$  cm, and rigidity  $\mu = 4 \times 10^{11}$  dyn cm $^{-2}$ , the seismic moment is  $\bar{M} = 1.0 \times 10^{27}$  dyn cm. The stress drop calculated using a strike-slip model is  $\Delta\sigma = 35$  bar. Comparing this value with those of other earthquakes, we find that the stress drop falls within the general range of those inter-plate earthquakes<sup>6</sup>. The observed displacements and calculated stress drop, however, are high as compared with some other strike-slip earthquakes. The relatively young age of the fault and the recent basalt flows across which the fault cuts may be responsible for this deviation.

There were no instruments in the area except for a seismoscope in Van, about 90 km to the south-west of the epicentre. A county centre, however, (Caldiran) and about 20 villages were located directly on the fault trace. The reports from these and other affected villagers were gathered (mostly through personal interviews) regarding the pre- and post-earthquake phenomena. The earthquake occurred at 2:22 p.m. local time. Most villagers raise livestock. About half the people were outdoors tending their herds. Because of high regional elevations, the weather was already cold and in the western end, the ground was covered with snow.

There were no foreshocks felt along the fault zone before the main shock on 24 November 1976. In two villages along the western half of the fault trace, where displacements were highest, noises resembling thunder were heard several times during the week preceding the quake. These could have been due to small foreshocks, even though the shaking was not felt. In the village of Hidirmentes, directly on the fault trace and where it cuts the lake, there were reports of noises coming from the lake (foreshocks, gas release, microfracturing?) during the same period. There was no observable change in water level before the earthquake.

There was one reported case of change in ground water activity before the earthquake: in the village of Yukarigulderen, near the eastern end of the fault trace, discharge from a spring increased. Also associated with this increase was seepage of oil that covered the water surface. There are cases of oil seepage in this general area.

After the earthquake, major changes in ground water conditions were noticed both near the fault trace and on the southern side. Some springs became muddy and stayed so for days. A few springs dried on the north side of the fault and a well went dry in the south. At Tuzla, 50 km south of the fault trace, discharge from the salt spring increased by a factor of 20 immediately after the earthquake and it was three times normal a week later. The north shore of Lake Van was uplifted relative to the south shore by about 16 cm. The southward tilting of the lake was most likely co-seismic, but a deformation before the earthquake cannot be ruled out.

Behaviour of domestic and farm animals before the earthquake was investigated extensively by interviewing many villagers and shepherds. In this primarily ranching area, there are cattle, sheep, goats, horses and dogs. The animals are housed in barns adjacent to the houses, although at the time of the earthquake most were outdoors. We inquired in widely scattered villages along the fault zone. There were no confirmed observations of unusual behaviour of farm animals, outdoors or indoors, before the earthquake. There was one report of animal restlessness in a village on the western end of the fault that reportedly lasted for months before and after the quake. This was not observed at neighbouring villages. The barking and howling



Fig. 2 Photographs of the fault trace observed immediately after the earthquake. *a*, The fault trace photographed from a distance. *b*, The right-handed displacement is clearly visible by the 250-cm offset of a ditch.

dogs were widely reported in a number of villages located on the west, central and eastern parts of the fault. The reported timing, however, of these events varied from a few minutes to a few hours before the earthquake.

Since there had been no major earthquakes in this region for a few generations, the villagers in general had no ideas or preconceptions about earthquakes. We believe that the reports we received were unbiased. There was no wide-scale abnormal behaviour of farm animals before the earthquake, except possibly for the behaviour of dogs. Physical precursors were definitely observed before the earthquake.

We thank the MTA Institute and its van regional office for support which made the field studies possible. Dr A. Boray contributed the regional map in Fig. 1. This research was supported by the Advanced Research Projects Agency, monitored by the Air Force Office of Scientific Research (contract F44620-75-C-0064) and by NATO (Science Grant No. 1162).

M. NAFI TOKSOZ

Department of Earth and Planetary Sciences,  
Massachusetts Institute of Technology,  
Cambridge, Massachusetts 02139

ESEN ARPAT  
FUAT ŞAROĞLU

MTA Institute,  
Ankara, Turkey

Received 8 August; accepted 19 October 1977.

1. Ketin, I. *Bull. Min. Res. Explor. Inst. Turkey, Ankara* **66**, 23–34 (1966).
2. Arpat, E. & Şaroğlu, F. *Bull. Min. Res. Explor. Inst. Turkey, Ankara* **78**, 33 (1972).
3. McKenzie, D. *Geophys. J. R. astr. Soc.* **30**, 109–185 (1972).
4. Ergin, K., Güçlü, U. & Uz, Z. *Technical Univ. Istanbul. Mining Eng. Fac., Publ. No. 24*, 74pp (1967).
5. Alsan, F., Tezcan, L. & Bath, M. *Tectonophysics* **31**, 13–20 (1976).
6. Berberian, M. *Geol. Survey of Iran, Report No. 39* (1976).
7. Arpat, E., Şaroğlu, F. & İz, H. B. *Yerleşim ve İmar*, **2** (1) 29–41 (1977). In Turkish.
8. Kanamori, H. & Anderson, D. L. *Bull. seism. Soc. Am.* **65**, 1073–1095 (1975).



## Timing of continental growth and emergence

HARGRAVES<sup>1</sup> has proposed that the separation of the primordial shells of the earth into continents and oceans began only about 3,700 Myr ago and continued through most of Precambrian time, with the result that the bulk of the continents as we know them today did not emerge from the sea until 1,400–1,000 Myr ago. Here I review geological and geochemical data that indicate that the growth and emergence of the bulk of the present-day continents took place in or by  $2,500 \pm 200$  Myr ago.

There are many estimates of the thickness that the continents had reached by the late Archaean. First, studies of mineral assemblages and chemistry correlated with experimental data on 3,000 Myr high-grade rocks from Greenland, Scotland, Africa and India suggest they formed at depths of 30–50 km (refs 2–5) and, because the Moho is 35–40 km beneath them, the Archaean crust in these regions was probably at least of the order of 65–80 km thick<sup>3,8</sup> (this discounts the unattractive alternative of subsequent sialic underplating to make up the difference), comparable to the continental crust today in the Andes<sup>9</sup>. These calculations are not surprising if the major period of crustal growth took place in the late Archaean by Cordilleran plate margin activity<sup>3,7</sup>. Extensive thickening of the crust in the late Archaean could have been effected by a combination of thrust-nappe stacking and magma addition<sup>8–10</sup>. Second, for greenstone belts mostly in North America, Condic<sup>2</sup> calculated crustal thicknesses primarily in the range of 20–30 km from Rb–Sr data, and of 20 km from K<sub>2</sub>O–SiO<sub>2</sub> data. Potash and Rb–Sr indices of granitic rocks suggest crustal thicknesses equal to or more than 30 km at the time of granite emplacement late in the evolution of the greenstone belts. Consideration of the spacing between volcanic centres in the Abitibi belt in relation to the known dependence of modern volcano spacings on crustal thickness gives an Archaean crustal thickness of 35–45 km (ref. 11). In comparison with the thickness estimates from the high-grade regions these values are consistent with formation of greenstone belts in some primitive form of marginal basin environment<sup>12,13</sup>. In spite of various problems connected with the interpretation of geobarometry, the above estimates suggest that the major stage of crustal fractionation was completed by the end of the Archaean.

There are many reasons for regarding the Archaean–Proterozoic time boundary (some  $2,500 \pm 200$  Myr ago) as the most important turning point in the evolution of the continents<sup>14–16</sup>; it represents the changeover from a permobile regime to a platform-geosynclinal style of tectonics that begins to show strong similarities with that of modern times. Indications of the new-found stability of the early Proterozoic continents include the intrusion of transcontinental basic dyke swarms, and the first appearance of alkaline complexes, kimberlites, carbonatites and aulacogens that extend from continental margins into the interior of stable plates. But the best evidence of the widespread emergence of stable cratons in the early Proterozoic is the appearance of miogeoclinal and geosynclines with the first recognisable continental margin sequences.

Many Archaean cratons are bordered, or partly underlain, by early Proterozoic sedimentary basins—an indication that the cratons were positive emergent areas bordered by shorelines in the early Proterozoic. One of the best-documented studies of craton-basin relations is by Hunter<sup>17</sup>, who has followed the development of basins in the Kaapvaal craton of southern Africa from about 3,300 to 1,850 Myr ago. He showed that the rates

of vertical movement of the basins decreased with time, which is consistent with the observed increase in volume of finer clastics and non-clastics in successively younger basins, and also estimated the percentage increase in inundation of the Kaapvaal craton by shallow Proterozoic seas, which took place by episodic regressions and transgressions. The data indicate that the craton was an emergent area more in the Archaean than the Proterozoic, the peak of continental emergence being reached between 3,350 and 3,000 Myr ago. Exposed mountains were gradually eroded during the Proterozoic, increasing the availability of sialic clastic debris and enabling increasing transgressions by epicontinental seas. This picture is at variance with the proposals that the Archaean continents were submerged beneath a globe-encircling sea, that the height of mountains increased during the Proterozoic, and that continental emergence did not take place on a substantial scale until after 1,400 Myr ago in the late Proterozoic<sup>1</sup>.

Table 1 Some early Proterozoic geosynclines and basins in the North American Shield

| Succession                          | Abbreviation on Fig. 1 | Thickness (km) | Ref. |
|-------------------------------------|------------------------|----------------|------|
| Huronian Supergroup                 | H                      | ~12.0          | 19   |
| Animikie Group                      | A                      | 7.5            | 19   |
| Coronation Geosyncline              | C                      | 10.7           | 20   |
| Circum-Ungava Geosyncline           | CU                     | >15.0          | 21   |
| Ramah Group, Labrador               | R                      | 1.7            | 22   |
| Baffin Geosyncline                  | B                      | ?              | 23   |
| Chibougamau Group                   | CB                     | 1.1            | 24   |
| Wollaston Lake Fold Belt            | W                      | ?              | 21   |
| Thompson Fold Belt                  | T                      | ?              | 21   |
| Hurwitz Group (Kaminak Subprovince) | K                      | 4.0            | 21   |
| Medicine Bow Mountains, Wyoming     | M                      | 8.0            | 24   |

Hunter's data are echoed in a more general way by Watson's<sup>18</sup> review of vertical movements during the Proterozoic. She concludes that erosion of Archaean cratons down to the present level of exposure was completed in periods of no more than 300 or 400 Myr after their stabilisation, and that the cratons typically underwent very limited vertical movements after about 2,000 Myr ago.

The North American shield contains a large number of early Proterozoic geosynclines and basins (in the Aphebian age range, 2,560–1,800 Myr), some of which are shown in Table 1 (see also Fig. 1).

These Aphebian successions, many of which are unconformable, overlie or border Archaean cratons for a minimum total length of 8,000 km (Fig. 1). Besides iron formations and dolomites the sediments consist predominantly of clastic conglomerates, quartzites, siltstones, shales and sandstones. It is evident that over the whole shield Archaean mountains were being eroded to provide great thicknesses of terrigenous clastic debris in the early Proterozoic<sup>14,25</sup>. The Huronian is a miogeoclinal sequence containing craton-derived sediments on the south side of the Superior Province<sup>19</sup>. The Coronation Geosyncline has a stratigraphy and structure that so faithfully mirrors that of modern Cordilleran successions that it is highly likely that modern-style plate tectonics began by the early Proterozoic<sup>20</sup>. The Circum-Ungava Geosyncline probably formed in a narrow ocean or rift that was closed by continent–continent collision<sup>26,27</sup>. The North American successions are no doubt only remnants that were once more extensive, but, nevertheless, they are sufficient indication that the bulk of the Archaean shield was undergoing uplift and emergence in the early Proterozoic.

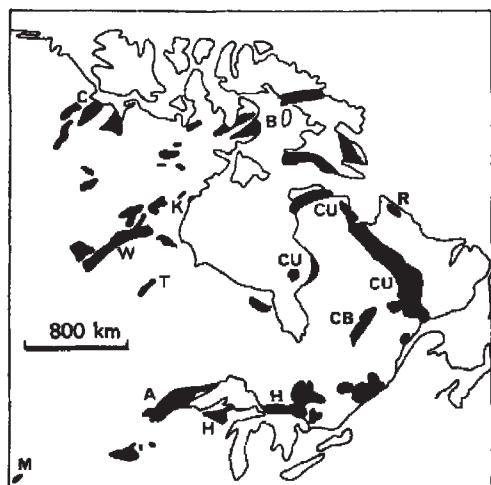


Fig. 1 Early Proterozoic geosynclines in the North American Shield. For abbreviations see Table 1. Compiled from refs 21, 24, 25.

There are well-known comparable early Proterozoic successions on all other continents, which abound in clastics and which indicate that Archaean cratons the world over were being highly eroded and many were at or near sea level by the early Proterozoic.

Rb-Sr and Pb-U isotope systematics yield information on the timing of fractionation of continental crust. For example, the growth of the strontium isotope ratio with time and the variable intercepts with the primary growth curve in the lead isotope system may be used to distinguish juvenile addition or accretion of mantle-derived material to the continental crust from rejuvenation or reworking of continental basement<sup>28,29</sup>. Table 2 shows estimates of the percentages of present-day continental crust formed by the late Archaean.

These values agree with estimates, based on areas of rock distribution, that at least 50% of North America<sup>34</sup> and Africa<sup>35</sup> were in existence by 2,500 Myr ago. If 50% of the volume of the present continental crust was formed by about 2,500 Myr ago it is possible that the continents have been growing more or less linearly during geological history<sup>33</sup>. There is increasing evidence, however, that the growth of the present-day continents, both in terms of surface area and thickness, was probably almost complete by 2,500 Myr ago (see below).

The model strontium isotope curve for the evolution of seawater shows a marked increase about 2,600 Myr ago, which is consistent with a major stage of crustal fractionation at that time<sup>32</sup>.

The factors of continental area and thickness, and the evidence of early (Archaean) crustal thickening and of early (late Archaean) emergence of continents above sea level, are interrelated. This interdependence is well illustrated by the conclusions of Wise<sup>36</sup> of freeboard (elevation of continents with respect to sea level) as a function of time: (1) the continents have had at least

90% of their present thickness for a minimum of 2,500 Myr. (2) There has been isostatic balance (constant freeboard) between oceanic and continental crusts throughout the Proterozoic and Phanerozoic, that is continental and oceanic volumes and areas have not changed significantly in this time span.

This picture is supported in particular by Watson's<sup>34</sup> conclusion about the lack of vertical movement of Archaean cratons for the last 2,000 Myr, and by the fact that the present thickness of continental crust in belts reworked at different times throughout the last 2,500 Myr is constant (38–40 km range) and independent of age<sup>37</sup>.

Isostatic modelling<sup>1</sup> shows that a 30-km thick crust can support ideally mountains that rise 2 km above sea level. Because there is widespread evidence that the continental crust had reached such a thickness by the late Archaean, it can be supposed that there were 2 km high mountains at that time in greenstone belt country—a likely result of vertical movement at a rate of 0.27 km Myr (ref. 17)—and even higher mountains above granulite-gneiss belts. Erosion of such elevated continents may account for the high proportion of immature sediments (arkoses and greywackes) in late Archaean greenstone belt sequences, the recycling of which during later geological time contributed to more mature detrital sediments (shales, sandstones and quartzites)<sup>38,39</sup>.

In Phanerozoic time continental fragmentation was accompanied by transgressions, and continental assembly by regressions<sup>40</sup>; it has long been known that the periods immediately following the formation of the Caledonian, Hercynian and Alpine fold belts were characterised by regressions<sup>38</sup>. In the mid-late Archaean the most intense period of 'orogenesis' and continental growth in earth history most likely resulted in the assembly of a large number of small plates (because of a greater ridge and subduction zone length) into more extensive, thick and stable plates by the early Proterozoic<sup>15,16</sup>. Because regression follows orogeny and accompanies continental assembly, the Archaean-Proterozoic boundary probably marks a period of profound regression followed, in the early Proterozoic, by widespread transgression that led to the formation of thick carbonate banks (for example, Coronation and Mount Isa Geosynclines) and miogeoclines (for example, Huronian Supergroup and the Western Labrador Trough).

BRIAN F. WINDLEY

Department of Geology,  
University of Leicester,  
Leicester, UK

Received 27 July; accepted 11 October 1977.

- Hargraves, R. B. *Science* **193**, 363–371 (1976).
- Wells, P. R. A. *Contr. Mineral. Petrol.* **56**, 229–242 (1976).
- Condie, K. C. *Plate Tectonics and Crustal Evolution* 288 (Pergamon, Oxford, 1976).
- O'Hara, M. J. *J. geol. Soc. Lond.* (in the press).
- Tamey, J. & Windley, B. F. *J. geol. Soc., Lond.* (in the press).
- James, D. E. *Bull. geol. Soc. Am.* **82**, 3323–3346 (1971).
- Windley, B. F. & Smith, J. V. *Nature* **260**, 671–675 (1976).
- Bridgwater, D., McGregor, V. R. & Myers, J. S. *Precamb. Res.* **1**, 179–197 (1974).
- Coward, M. L., Lintern, B. C. & Wright, L. in *The Early History of the Earth* (ed. Windley, B. F.) 323–330 (Wiley, London, 1976).
- England, P. C. & Richardson, S. W. *J. geol. Soc., Lond.* (in the press).
- Windley, B. F. & Davies, F. B. (in preparation).
- Tamey, J., Dalziel, I. W. D. & de Wit, M. J. in *The Early History of the Earth* (ed. Windley, B. F.) 131–146 (Wiley, London, 1976).
- Burke, K., Dewey, J. F. & Kidd, W. S. F. in *The Early History of the Earth* (ed. Windley, B. F.) 113–129 (Wiley, London, 1976).
- Salop, I. *Precambrian of the Northern Hemisphere* 378 (Elsevier, Amsterdam, 1977).
- Burke, K. C. A. & Dewey, J. F. in *Implications of Continental Drift to the Earth Sciences* (eds Tarling, D. H. & Runcorn, S. K.) 2, 1035–1045 (Academic, London, 1973).
- Windley, B. F. *The Evolving Continents* 385 (Wiley, London, 1977).
- Hunter, D. R. *Precamb. Res.* **1**, 295–326 (1974).
- Watson, J. V. *Phil. Trans. R. Soc. A280*, 629–640 (1976).
- Card, K. D. et al. *Geol. Ass. Can. Spec. Pap.* **11**, 335–380 (1972).
- Hoffman, P. *Phil. Trans. R. Soc. A273*, 547–581 (1973).
- Davidson, A. *Geol. Ass. Can. Spec. Pap.* **11**, 381–434 (1972).
- Morgan, W. C. *Geol. Surv. Can. Pap.* **74–54**, 42 (1975).
- Jackson, G. D. & Taylor, F. C. *J. Can. Earth Sci.* **9**, 1650–1669 (1972).
- Young, G. M. *Palaogeog. Palaochim. Palaecol.* **7**, 85–101 (1970).

Table 2 Estimated percentages of continental crust formed by the late Archaean

| Isotopes | % of present-day continents | Time by which % was formed | Ref. |
|----------|-----------------------------|----------------------------|------|
| Pb-U     | Most                        | 2,500                      | 30   |
| Pb & Sr  | Almost all                  | 2,500                      | 31   |
| Rb-Sr    | Almost all                  | 2,700–2,500                | 32   |
| Rb-Sr    | 67                          | 2,700–1,800                | 33   |
| Rb-Sr    | 50                          | 2,900–2,500                | 28   |



25. Douglas, R. J. W. & Price, R. A. *Geol. Ass. Can. Spec. Pap.* 11, 625-688 (1972).
26. Gibb, R. A. *Earth planet. Sci. Lett.* 27, 378-382 (1975).
27. Kearey, P. *Earth planet. Sci. Lett.* 28, 371-378 (1976).
28. Moorbath, S. *Proc. Geol. Ass.* 86, 259-279 (1975).
29. Moorbath, S., Welke, H. & Gale, N. H. *Earth planet. Sci. Lett.* 6, 245-256 (1969).
30. Patterson, C. C. & Tatsumoto, M. *Geochim. cosmochim. Acta* 28, 1-22 (1964).
31. Armstrong, R. L. *Rev. Geophys.* 6, 175-199 (1968).
32. Jahn, B. M. & Nyquist, L. E. in *The Early History of the Earth* (ed. Windley, B. F.) 55-76 (Wiley, London, 1976).
34. Muehlberger, W. R., Denison, R. E. & Lidiak, E. G. *Am. Ass. Petrol. Geol. Bull.* 51, 2351-2380 (1967).
35. Kroner, A. *Precamb. Res.* 4, 163-214 (1977).
36. Wise, D. U. in *The Geology of Continental Margins* (eds Burk, C. A. & Drake, C. L.) 45-58 (Springer, Berlin, 1974).
37. Condie, K. C. *Bull. geol. Soc. Am.* 84, 2981-2992 (1973).
38. Ronov, A. B. *Geochem. Int.* 1, 713-737 (1964).
39. Veizer, J. *Contr. Mineral. Petrol.* 38, 261-278 (1973).
40. Valentine, J. W. & Moores, E. M. *Nature* 228, 657-659 (1970).

## Alcohol retards visual recovery from glare by hampering target acquisition

ADAMS and Brown<sup>1</sup> showed that relatively small doses of alcohol produced "large, significant, dose-related increases in the time required to recover foveal contrast sensitivity following bright light exposure". As this effect of alcohol may be directly related to industrial and to car driving safety<sup>2</sup>, we set out to ascertain its precise origin. Adams and Brown concluded tentatively that the alcohol exerted its effect at the retinal level. The experiments presented here, however, implicate non-retinal mechanisms, as alcohol delays recovery from glare only when observers have difficulty localising or fixating the test stimulus.

The first part of our study was a replication of the Adams and Brown experiment. Our subjects were four male college students with normal (6/6) or corrected-to-normal visual acuity. On separate days subjects ingested either alcohol (1.0 ml per kg body weight) or placebo. The two treatments were given in random order. The alcohol dose was 95% ethanol diluted 2:1 with fruit drink so that the total volume (ml) was three times the subject's weight (kg). The placebo was an equivalent volume of fruit drink alone. The drink was consumed in 20 min from a paper cup containing two ice cubes. In addition, 2 drops each of ethanol and eucalyptus oil were placed on the lid of the cup to minimise olfactory cues to the presence of alcohol<sup>1</sup>.

Details of stimulus presentation are given in Fig. 1. Subjects practiced the detection task to achieve stable performance at the beginning of each experimental session; they then were tested before drinking, and 30, 75, 150, 240, and 330 min after drinking. At each of these times, three sets of measurements were taken with 5 min allowed between sets to allow complete recovery from the adapting luminance level. The session always started in the early afternoon, and subjects were instructed to eat lunch beforehand.

As expected, alcohol produced large, significant increases in the time required to recover foveal contrast sensitivity. In particular, as Fig. 1 shows, the difference between recovery times with and without alcohol is maximum at the sampling time 75 min after ingestion and declines thereafter, and is most pronounced at low target contrasts. Analysis of variance (ANOVA) indicates that the depicted interaction between dose, time, and contrast level is statistically significant ( $P < 0.001$ , d.f. = 20, 60).

A simple but trivial explanation of the results of Adams and Brown and of our replication would be that following alcohol ingestion observers become more cautious about reporting the visibility of the test target, thereby spuriously increasing recovery times<sup>3,4</sup>. We decided therefore to use a methodology<sup>5</sup> which could discriminate alcohol-induced changes in sensitivity from changes in criterion (that is, willingness to report).

Using the same apparatus as before but without a glare source, we tested seven male college students. On half of the 2-s trials the small flickering test spot actually was presented; on the remaining trials no spot was presented. The subject's task was to respond either 'yes' (a test spot had been presented) or 'no'. With each subject, 500 trials were run at various spot

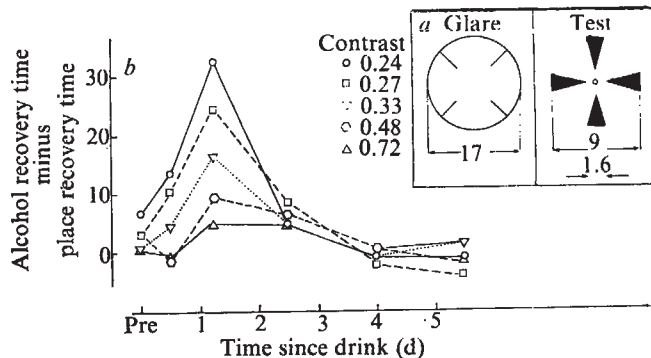


Fig. 1 a, The stimulus configuration of the  $8.6 \times 10^4$  cd m<sup>-2</sup> glare field on the left and the test field on the right. The 10 min arc test spot was located at the centre of the test field reference markers on a background of luminance 24.8 cd m<sup>-2</sup>. In the actual experiment, the glare field was located directly above the test field, which was at eye level. At the beginning of each measurement session, the subject fixated with his left eye the centre of the circular adapting (glare) field. The right eye was covered with a patch, and fixation was aided by four thin diagonally-orientated reference lines. Immediately after a 60-s exposure to the high intensity glare source, and with the eye-patch still in place, the subject looked directly below (straight ahead) and fixated the centre of the test field on which a test spot was presented intermittently (125-ms flashes at 4 Hz). The contrast of the test spot was under the experimenter's control. When the subject indicated orally that he saw the spot at the highest contrast level, the experimenter rotated a filter wheel one notch, reducing the target contrast a further step below the subject's threshold. When the subject's contrast sensitivity had recovered enough to restore the test spot to detectability, the spot contrast was reduced again. The time taken to recover to each of five predetermined contrast levels (0.72, 0.48, 0.33, 0.27, and 0.24) was obtained. These levels were chosen to give approximately equal intervals between each recovery point. b, Interaction of alcohol, contrast, and time for the first experiment. Each point is the mean difference (s) between recovery time with alcohol and recovery time with the placebo. Data from four subjects.

contrasts. Subjects were tested either following ingestion of alcohol (dose identical to that used before) or following ingestion of the placebo. Intermittent tests were carried out 30-140 min after ingestion, when glare recovery was most seriously retarded in our original experiment. The conditional probabilities of 'yes' responses in the presence of a test spot and 'yes' responses in the absence of a test spot were used to estimate<sup>6</sup> the observer's response criterion ( $\beta$ ). Analysis of variance indicated no significant alcohol-induced changes in  $\beta$  ( $P > 0.50$ ).

With response bias eliminated as an explanation, we sought to identify the source of the alcohol-induced delay in glare recovery times more precisely. Four new subjects participated in an experiment identical in every respect to our first, except that three sets of eye movement measurements (before, during, and after exposure to the glare) were taken every 15 min for 135 min following the 20-min drinking period. Standard electrooculographic<sup>6</sup> (EOG) methods produced records of eye position as a function of time. The resolution of our recordings was approximately one degree of visual angle.

With or without alcohol, very little eye movement occurred either before or during exposure to the glare source. But immediately after the 60 s of high intensity glare, there were many eye movements as if subjects 'hunted' for the flashing spot within the test field; after alcohol ingestion the eye movements were larger and more frequent. These eye movement patterns suggested that alcohol might combine with glare to prolong recovery time by impeding the subject's attempts to localise and fixate the target. Localisation would be hampered if alcohol ingestion were to disturb the subject's memory of precisely where the target would appear. Steady fixation would be hampered if alcohol were to impair control of the extraocular muscles<sup>7</sup>. Irregular visual fixation could produce enough



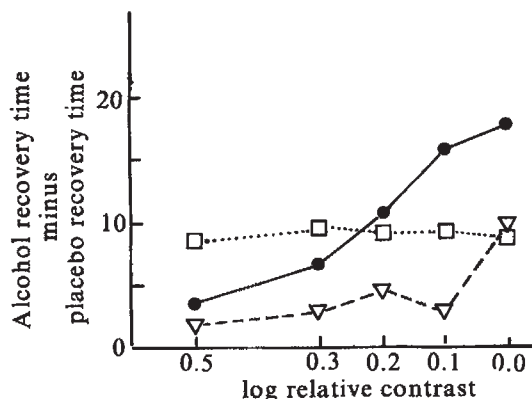
uncertainty about the target's spatial location to reduce its detectability<sup>8</sup>.

Three of the four subjects who participated in our original study were available for two additional experiments. In the first, the test spot locus was circumscribed by a small, dark circle, 12.5-arc min diameter. The circle was of high contrast, and clearly visible immediately after glare exposure even though the flashing test spot within was not. As a result, the subject would not have to remember where the target might appear thereby aiding target localisation. Details of the experimental design and procedure were as in the original experiment, except that the test spot luminance was three times that of the spot in the original experiment. In a second experiment, the test spot was enlarged (100-arc min diameter) to fill the entire area between the arrow reference marks (Fig. 1). Consequently, no matter where between the arrows the subject looked, his fixation would fall on some part of the flashing test spot. Enlarging the area of the test spot reduced its maximum luminance; to compensate, we reduced the background luminance to 0.89 cd m<sup>-2</sup>. In both these experiments, as before, glare recovery was measured with spots whose contrasts covered a range of 0.5 log units; a single set of five measurements was made before, and 30, 90, 150 and 210 min after drink ingestion.

As shown by Fig. 2, the effect of alcohol on recovery from glare was greatly reduced when either the small test spot was circumscribed by a clearly defined circle or the test spot was enlarged. With the enlarged test spot, no effects involving alcohol were statistically significant (ANOVA, all  $P > 0.5$ ). This is what would be expected if the alcohol-induced retardation of glare recovery in our original experiment had been partly caused by an unsteadiness of fixation which the enlarged test spot rendered unimportant. The contrast independent elevation shown in Fig. 2 is not statistically significant ( $P > 0.80$ ) and is due to greatly increased recovery times for just one subject. With the small, circumscribed test spot, one effect involving alcohol (the interaction of alcohol and contrast) was significant ( $P < 0.03$ , d.f. = 4, 8), due mainly to the increased recovery time at the lowest contrast level. This reduced effect of alcohol would be expected if in our original experiment alcohol had made subjects unsure about the location of the target and if the easily seen, high-contrast circle in the present experiment had made them less unsure. The small residual effect of alcohol may reflect difficulty in maintaining precise fixation following alcohol consumption, a problem that would be less critical with the enlarged spot.

Our experiments confirm that ingestion of alcohol causes a

**Fig. 2** Mean difference (s) between recovery time with alcohol and recovery time with the placebo. Data are shown for all five contrast levels and for all three glare experiments. The data have been normalised in both this figure and in all analyses of variance to produce 47.52-s average recovery time for each subject (the average for the original experiment). Only the data from those subjects who participated in all three experiments are shown. Data shown were collected at those times following alcohol ingestion which should have produced the largest effect of alcohol: 75 min for the original experiment (●), 90 min for the large spot (□), and 90 min for the circumscribed spot (△).



loss of visual sensitivity following glare. But non-retinal processes are primarily responsible for the effect of alcohol on glare recovery. We emphasise that the non-retinal origin of this alcohol-induced visual disability does not lessen its potential for adversely affecting driving and related visual tasks. On the contrary, uncertainty about the location<sup>8</sup> of possible targets imposes an inescapable reduction of visual sensitivity of any driver; if alcohol either retards a driver's target acquisition<sup>9</sup> or causes him to mislocalise the target<sup>10</sup>, visibility necessarily will be reduced and his ability to respond to the target impaired. Moreover, we cannot rule out the possibility that in driving, the effect of alcohol that we studied can sum with other, previously identified perceptual consequences of alcohol ingestion<sup>11</sup>.

This research was supported by grants from the National Eye Institute (EY-00321), the NSF (BNS77-15858) and Northwestern University Research Committee. We thank Peter Herman and Tony Adams for useful discussions and Randolph Blake for constructive comments on a draft of this manuscript.

ROBERT SEKULER

RODGER D. MACARTHUR

*Cresap Neuroscience Laboratory,  
Department of Psychology,  
Northwestern University,  
Evanston, Illinois 60210*

Received 27 June; accepted 3 October 1977.

- Adams, A. J. & Brown, B. *Nature* 257, 481-483 (1975).
- Moskowitz, H. & Murray, J. T. *J. Stud. Alcohol* 37, 40-45 (1976).
- Tong, J. E., Knott, V. J., McGraw, D. J. & Leigh, G. Q. *J. Stud. Alcohol* 35, 1003-1022 (1974).
- von Wright, J. M. & Mikkonen, V. *Scand. J. Psychol.* 11, 167-175 (1970).
- Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
- Young, L. R. & Sheena, D. *Behav. Res. Meth. Instr.* 7, 397-429 (1975).
- Wilkinson, I. M. S., Kime, R. & Purnell, M. *Brain* 97, 785-792 (1974).
- Cohn, T. E. & Lasley, D. J. *J. opt. Soc. Am.* 64, 1715-1719 (1974).
- Adams, A. J., Brown, B., Flom, M. C., Jones, R. T. & Jampolsky, A. *Am. J. Optom. Physiol.* 52, 729-735 (1975).
- Greenhouse, D. S., Cohn, T. E. & Stark, L. presented at *Mtg Ass. Res. Vis. Ophth.* Sarasota, Florida (April 1977).
- Moskowitz, H. & Sharma, S. *Human Factors* 16, 174-180 (1975).

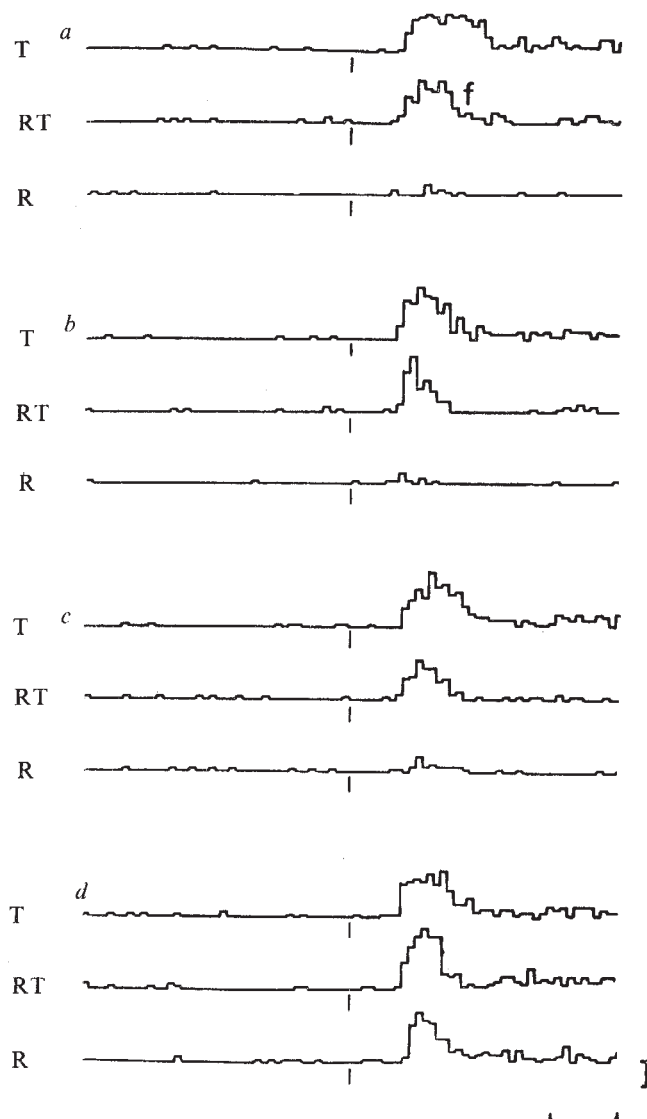
## Visual discrimination between small objects and large textured backgrounds

THE remarkable ability of certain higher order insect visual neurones to discriminate between movement of a small object or target and a large textured background has been clearly demonstrated by Palka<sup>1,2</sup> and O'Shea and Fraser-Rowell<sup>3</sup>. In addition, movement of a large textured background such as a windmill pattern is inhibitory to detection of motion of a small target by this neurone, known as the descending contralateral movement detector (DCMD). The response of this identifiable neurone to motion of small objects is inhibited by rotation of the windmill pattern over a specific range of high spatial frequencies of the pattern, but is augmented by low spatial frequencies in the pattern. For this neurone we have quantitatively determined the spatial frequency at which the effect of the windmill pattern becomes inhibitory.

The DCMD is a visual interneurone in the ventral nerve cord which is excitatory to motor neurones involved in jumping. It originates in the protocerebrum where it receives excitatory input through an electrical synapse from the LGMD (lobular giant movement detector) which has a dendritic field in the shape of a fan across the entire lobula. The LGMD receives at least two kinds of inhibitory inputs: lateral inhibition on neurones peripheral to the fan of dendrites, and inhibitory input proximal to the convergence of the fan of dendrites (post-convergence inhibition)<sup>3,4</sup>. The post-convergence inhibition samples activity across the entire visual field but is effective only for higher velocities<sup>3,4</sup> of patterns than used in these experiments.

Responses of the DCMD in *Schistocerca gregaria* were obtained by conventional methods<sup>3</sup>. The receptive field of the DCMD is approximately a hemisphere, and motion detection

does not depend on the direction of motion<sup>5,6</sup>. The region of greatest sensitivity lies along the equator posterior to the animal's transverse plane, and it is primarily this region which was tested in these experiments. Two visual stimuli were presented. One, an approximately 5° (in visual angle subtended at the cornea) light disk which is occluded by a moving edge, yields a large sustained train of spikes over the entire time



**Fig. 1** Sets of PST histograms demonstrating inhibition of small target response by radial grating pattern. This set of histograms is to be viewed in sequential groups of three traces. R refers to rotation of the radial grating pattern only, T refers to motion of the edge in the small disk only, and RT refers to simultaneous grating pattern rotation and edge motion. Radial gratings are of 90 black and white stripes of equal angular width (a), 40 stripes (b), 20 stripes (c) and 10 stripes (d). This figure shows, for 90, 40 and 20 stripes, that the excitatory response to the motion of the small target alone (traces T) precedes the inhibition due to the simultaneous radial grating rotation (traces RT). In the second trace from the top (a), marks inhibition occurring at the falloff of response to small target motion. Patterns and disk are adjacent along the equator with disk posterior, and are 9 inches from the cornea; the radial grating subtends 37° visual angle and the disk 5.6°. Time of initiation of motion is shown by small vertical bar near centre, below each trace, so that spontaneous activity is seen to the left; motion terminates at end of record. Each trace represents 17 trials, and the T and RT are interleaved with 30 s interstimulus interval (ISI). The R are a separate run with 30 s ISI. Experiment of 21 February 1976, light surround (see Fig. 2); order of T/RT was 90, 20, 10, 40 stripes; order of R was 10, 40, 20, 90; time bar 100 ms; vertical scale 59 spikes per s; sustained baseline is 0 spikes per s.

course of motion of the edge. The other, a large circular radial grating pattern (windmill pattern) of an adjustable number of alternating black and white stripes of equal width which is rotated about its centre, yields a burst of spikes only on initiation of rotation regardless of velocity.

In Fig. 1 the post-stimulus time (PST) histograms are shown for a typical experiment to illustrate the nature of the inhibition in time. For these experiments the clearest distinction between small objects and large backgrounds can be made in the spatial frequency domain. The spatial frequency spectrum of an intensity pattern is given by the Fourier transform of that pattern. This calculation shows that the magnitude of the spatial frequency spectrum of the radial grating pattern has a minimum sensible spatial frequency of approximately  $0.85F$  and peak spectrum spatial frequency of  $1.09F$ , where  $F$  is the single number spatial frequency geometrically calculated as  $1/L$ ,  $L$  being the angular distance subtended by one stripe cycle at the outermost edge of the pattern. In Fig. 2 this geometric value of  $F$  has been used, which represents the 25% of peak spectrum point. (Note: in ref. 7, page 336, the value  $0.85F$  erroneously reads  $1.3F$ , and  $1.09F$  erroneously reads  $1.64F$ .)

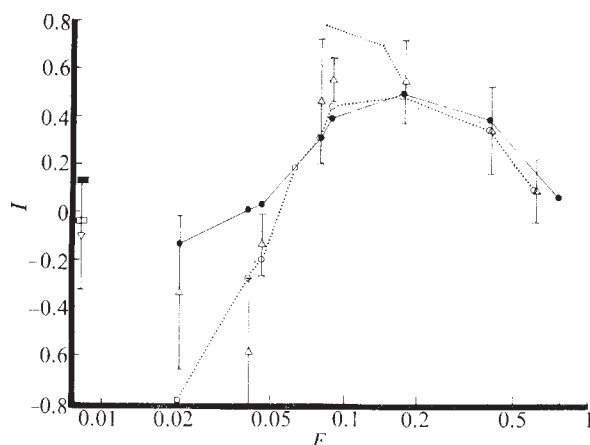
The spatial frequency components of the moving edge in the small disk are predominantly at low spatial frequencies, which is complementary to the situation for the radial grating pattern.

In Fig. 2 the inhibition  $I$  of the response to the small target caused by the radial grating pattern is plotted as a function of  $F$ , the nominal minimum spatial frequency of the radial grating. Since the curve is positive for  $F$  greater than 0.05 cycles per degree the radial grating is inhibitory when all of its spatial frequency components lie above 0.05 cycles per degree. When any of its components lie below 0.05 cycles per degree the radial grating becomes excitatory to the response to the small target, as the curve there is negative. The peak of the curve lies at about 0.2 cycles per degree which is thus approximately the spatial frequency of maximum inhibition. The curve falls to zero at 1.0 cycles per degree because that is the maximum resolvable spatial frequency for this animal<sup>7</sup>. Thus low spatial frequencies are excitatory and high spatial frequencies are inhibitory to the response to low spatial frequencies, as long as the high spatial frequencies are resolved. In Fig. 2 a further difference in the DCMD responses elicited by high and low spatial frequency stimuli is greater variability to low excitatory spatial frequencies than to high ones. At lower spatial frequencies fewer ommatidia are stimulated by a moving pattern so that higher integrative units have either fewer inputs to average, or are more susceptible to their inherent noise.

A peaked inhibition curve like that of Fig. 2 can be obtained with radial grating pattern diameters down to a minimum of 30° (visual angle). Inhibition could occasionally be found at 0.2 cycles per degree for radial grating pattern diameters down to a minimum of 10°. The conclusion is that the radial grating patterns become inhibitory backgrounds at the optimal inhibitory spatial frequency if their diameter exceeds 10°, but for inhibition to occur over a measurable range of spatial frequency their diameter must exceed 30°. The inhibition is not uniformly dependent on the distance between the radial grating pattern and the small target. The inhibition is also not dependent on grating rotation velocity below about 300° (pattern plane angle) per s.

A model for the observation that stimuli with predominantly high spatial frequency components are inhibitory to those with low ones can be based on optic lobe electrophysiology any neuroanatomy of locust<sup>8,9</sup> and dipterans<sup>10,11</sup>. Arnett<sup>10</sup> had shown that movement of a small contrasting pattern can be detected by first order interneurons (monopolar cells), and lateral inhibition has been demonstrated in locust optic lobe by Horridge *et al.*<sup>8</sup>, and O'Shea and Fraser-Rowell<sup>3,4</sup>. A detector of high spatial frequencies can be derived by spatial summation over the entire visual field of monopolar cell outputs, and a detector of primarily low spatial frequencies can be derived by spatial summation over the entire field of another set of monopolar cells having lateral inhibitory connections among





**Fig. 2** Inhibition ( $I$ ) of response to motion in small disk plotted as a function of the lower limit,  $F$  (cycles per degree visual angle subtended at cornea), of spatial frequency content of the radial grating pattern. Inhibition  $I$  is defined as  $(NE - NI)/NE$ , where  $NE$  is the number of spikes in the DCMD in response to motion in small disk alone, and  $NI$  is the number of spikes in response to simultaneous motion in disk and rotation of the radial grating pattern. When the number of spikes in response to simultaneous motion  $NI$  is greater than  $NE$ , inhibition  $I$  becomes negative and is, in fact, excitation. In all experiments, these two kinds of trials are interleaved to increase stability of the measurement. ●, Average of means of continuous series of 14 experiments 12 December 1975–14 February 1976 in completely dark surround. ○, Average of means of continuous series of three experiments in light surround, 17–21 February 1976. Light continuous dotted line, negative of response to rotation only of radial gratings from data of acuity studies of 1972<sup>7</sup> to show acuity limits inhibition at high frequencies. △, Data of one experiment (21 February 1976) in light surround; bars indicate the sum of standard error of  $NE$  and  $NI$ ; each point is the mean of 34 or 68 trials with 30-s ISI and interleaved presentation of disk motion trials with simultaneous disk motion and radial grating rotation. ■, Average of active controls on dark surround experiments. □, Average of active controls on light surround experiments. ▽, Active control on single light surround experiment. (An active control is rotation of the unilluminated radial grating with disk motion.) Stimulus parameters: luminance of a brighter bar of radial grating,  $1,200 \text{ lm m}^{-2}$ ; dark bar,  $\sim 0 \text{ lm m}^{-2}$ ; small disk has luminance equivalent to bright bar of radial grating; light surround luminance approximately  $1.5 \log_{10}$  units below that of bright bar; when the disk is occluded in light surround, the occluded portion has the luminance of the surround, but the excitatory response is approximately the same as in dark surround. The radial grating tangential velocity is  $11^\circ$  (visual angle) per s for most runs of this figure, or  $36^\circ$  per s in pattern plane angle. The velocity of translation of the occluding edge is  $12^\circ$  (visual angle) per s, and the light disk subtends  $5.6^\circ$  (visual angle); the radial grating subtends  $37^\circ$  (visual angle) for all but a few runs of this figure. In all experiments except as noted the disk and grating pattern are directly adjacent along the equator.

themselves. In the locust the latter summation would take place as excitatory synapses at or before the fan of LGMD dendrites in the lobula<sup>1</sup> via interneurons from the medulla, and the former summation as inhibitory synapses from the first set of monopolar cells on to these interneurons in the medulla. Since the LGMD drives the DCMD directly, this model gives for DCMD response two separate channels selectively responding to high and low spatial frequencies respectively, with an inhibitory cross-connection from the high frequency channel to the low frequency channel. Further tests of these ideas require probing of more peripheral units than the LGMD.

In vertebrate vision the reduction of retinal sensitivity during relative motion of retina and image is known as saccadic suppression. The suppressive or inhibitory effect on human vision (measured with a low spatial frequency test flash) produced by relative motion of gratings<sup>12,13</sup> is parallel to the above results for the locust DCMD. It is of interest that the locust visual system processes high and low spatial frequency image information in the same manner as do some perceptual

channels of the human visual system.

This study was supported by a grant (EY-01285) from the National Eye Institute of the NIH.

ROBERT B. PINTER

Departments of Electrical  
Engineering and Zoology,  
University of Washington (FT-10),  
Seattle, Washington 98195

Received 11 April; accepted 3 October 1977.

1. Palka, J. J. *exp. Biol.* **50**, 723–732 (1969).
2. Palka, J. *Am. Zool.* **12**, 497–505 (1972).
3. O'Shea, M. & Fraser-Rowell, C. H. *Nature* **254**, 53–55 (1975).
4. Fraser-Rowell, C. H., O'Shea, M., Williams, J. L. D. *J. exp. Biol.* **68**, 157–185 (1977).
5. Palka, J. J. *Insect Physiol.* **13**, 235–248 (1967).
6. Kien, J. in *The Compound Eye and Vision of Insects* (ed. Horridge, G. A.) 410–422 (Oxford University Press, Oxford, 1975).
7. Palka, J. & Pinter, R. B. in *The Compound Eye and Vision of Insects* (ed. Horridge, G. A.) 321–337 (Oxford University Press, Oxford, 1975).
8. Horridge, G. A., Scholes, J. H., Shaw, S. & Tunstall, J. in *The Physiology of the Insect Central Nervous System* (eds Treherne, J. E. & Beament, J. W. L.) 165–202 (Academic, London, 1965).
9. Meinertzhagen, I. *Phil. Trans. R. Soc. B*, **274**, 555–596 (1976).
10. Arnett, D. W. *J. Neurophys.* **35**, 429–444 (1972).
11. Strausfeld, N. J. & Campos-Ortega, J. A. *Science* **195**, 894–897 (1977).
12. Brooks, B. A. & Fuchs, A. F. *Vision Res.* **15**, 1389–1398 (1975).
13. Mateef, S., Yakimoff, N. & Mitrani, L. *Vision Res.* **16**, 489–492 (1976).

## Contribution of stored pre-anthesis assimilate to grain yield in wheat and barley

RESERVES of assimilate present in wheat and barley crops at flowering, and available for later translocation to the grains, could buffer grain yield against environmental stresses during grain filling. This so-called pre-anthesis assimilate contribution to grain yield can be expressed as a percentage of yield ( $P_1$ ). Archbold<sup>1</sup>, and later Thorne<sup>2</sup>, concluded that  $P_1$  was small, being no more than 20%.<sup>2</sup> But only one result (12% for irrigated wheat at Cambridge<sup>3</sup>) refers to a crop in the field as distinct from plants in pots, and no studies considered the effect of stress during grain filling. Recently Gallagher *et al.*<sup>4,5</sup> reported substantial contributions:  $P_1$  averaged 43% over six crops of wheat and barley at Nottingham; this amounted to more than 300 g per m<sup>2</sup> of dry material in two crops and, in the severe drought of 1970, 39% of total dry matter present at anthesis. They assumed, with some supporting evidence from one barley crop<sup>6</sup>, that the pre-anthesis contribution was given by the decrease from anthesis to maturity in dry weight of non-grain parts of the crop. *In situ* labelling with <sup>14</sup>CO<sub>2</sub> of the whole crop canopy at frequent intervals before and after anthesis would seem to be the least equivocal way of estimating  $P_1$ . Using this method we have determined  $P_1$  in wheat and barley. It averaged only 12% (watered crops) and 22% (droughted crops), and did not agree with estimates for the same crops obtained by the method of Gallagher *et al.*<sup>4,5</sup>.

Wheat and barley were grown during 1974–75 at the Centro de Investigaciones Agrícolas del Noroeste (CIANO) near Ciudad Obregon in north-west Mexico. There was no rain and crops were either flood irrigated frequently (I) or only until 27th January, about 40 d before anthesis, creating a drought (D) regime. Weather and general experimental procedures were as described elsewhere<sup>7</sup>. Grain yields for regime I were high (Table 1). Significant differences between D and I in leaf water potential arose about 10 d before anthesis and then increased steadily. There were substantial reductions in yield in regime D, especially for the wheat cultivars, which were later in flowering (Table 1).

The canopy labelling with <sup>14</sup>CO<sub>2</sub> began on 20 February and was repeated on separate 1.08 m<sup>2</sup> portions of the crops every 7–14 d until near maturity. The proportion of total above ground <sup>14</sup>C at maturity found in the grain was slightly greater for any given date of labelling in regime D, but was remarkably consistent between cultivars and species (Fig. 1). The relationship of Fig. 1a agrees with results<sup>2,8,9</sup> obtained in non-crop conditions.

Combining the data of Fig. 1 with information on changes in above ground dry weight,  $P_1$  was estimated to be 12–13% under



frequent irrigation, rising to 27% for droughted wheat (Table 2). But in absolute terms (grain dry weight,  $GDW_1$  in  $g\ m^{-2}$ ), it was similar for regimes I and D. As a proportion of total dry weight at anthesis,  $GDW_1$  amounted to 8.5% (I) and 11.5% (D). Considering that  $GDW_1$  for barley included the husk, already present at anthesis and amounting to 8% (I) and 9% (D) of final grain weight, the true contribution of stored pre-anthesis assimilate was clearly lower for barley than for wheat.

The method used to estimate  $P_1$  in Table 2 assumes that any contribution of carbon from roots, or variation in the proportion of carbon to dry matter in the total crop compared with that in the grain, or fluctuations in crop growth rate within the two key periods, are of minor importance. On the other hand, neglect of the loss of tissue, formed before anthesis, and lost during the post-anthesis period could be important. Such losses seem to be substantial under irrigation, probably due to the saprophytic decay of lower early-formed, and hence unlabelled, leaves. A loss of  $100\ g\ m^{-2}$  would seem probable (R.A.F., unpublished) under the I regime, and correcting  $\Delta TDW_2$  ( $TDW$ , total dry weight) for this would increase  $GDW_2$  by  $75\ g\ m^{-2}$  and lower mean  $P_1$  to 11%. We have also assumed that no assimilate formed before the date of first labelling was later transferred to the grain. Inspection of Fig. 1 shows this to be incorrect, but the small total dry weights at first labelling mean the amounts of assimilate neglected are small. The transfer to the grain of assimilate formed as early as 27 d before anthesis (Fig. 1a) may be associated with the transfer of leaf nitrogen in the form of protein.

Considering this correction to  $GDW_2$  under regime I, the values of  $GDW_1 + GDW_2$  agree closely with the observed grain yield (Table 2). Also the figure of 13% for irrigated wheat is similar to that of 12% for an irrigated wheat crop determined with  $^{14}CO_2$  by Lupton<sup>3</sup>, and to values which can be calculated with our procedure from the extensive results of Birecka and Dakic-Wlodkowska<sup>8</sup>, albeit obtained with wheat grown in pots. An independent upper estimate of the contribution of pre-anthesis assimilate in the case of adjacent irrigated crops of Yecora 70 at CIANO is given by the quantity of sugar stored in stems at anthesis. This amounted to  $50\ g\ m^{-2}$  (15% stem sugar;  $320\ g\ m^{-2}$  stem tissue), to which should be added leaf protein transfer to the grain, for these purposes set equal to the decrease, anthesis to maturity, in leaf lamina dry weight, amounting to  $40\ g\ m^{-2}$  ( $200\ g\ m^{-2}$  of green leaf lamina tissue, 20% decrease in dry weight excluding leaf loss). The total available material thus estimated ( $90\ g\ m^{-2}$ ) agrees well with the value of  $GDW_1$  of Table 2 ( $69\ g\ m^{-2}$ ).

Table 2 makes possible a separate estimate of  $P_1$  based on the decline in non-grain dry weight after anthesis ( $\Delta TDW_2$  less

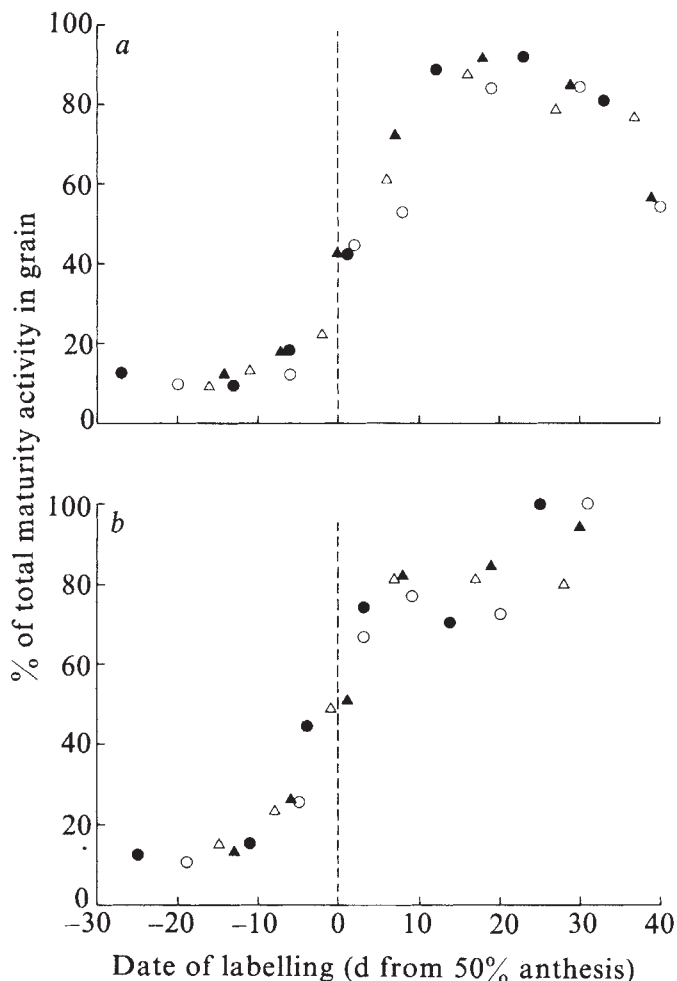


Fig. 1 The fraction at maturity of total crop  $^{14}C$  found in the grain as a function of the time (d from anthesis) when the  $^{14}CO_2$  was assimilated; a, frequently watered, (I); b, droughted, (D); Yecora 70 (●), Ciano (▲), CM67 (○) and WI 2198 (△).  $^{14}CO_2$  assimilation involved placing a clear polyethylene chamber over  $1.08\ m^2$  of crop. Into the chamber  $100\ \mu Ci$  of  $^{14}CO_2$  was introduced. The chamber remained closed and was stirred with a fan for 10 min. Labelling was carried out within 3 h of noon. At maturity, a  $0.3\ m^2$  quadrat was cut from each region previously labelled and the specific activity of oven dried subsamples were determined by combustion, trapping of the  $^{14}CO_2$  in a liquid scintillation mixture and counting.

Table 1 Grain yield and yield components in the 1974–75 experiment at CIANO, Mexico

| Species                            | Cultivar  | Date of 50% anthesis | Grain yield ( $g\ m^{-2}$ ) | Kernel weight (mg) |
|------------------------------------|-----------|----------------------|-----------------------------|--------------------|
| Frequently watered throughout (I)  |           |                      |                             |                    |
| Wheat                              | Yecora 70 | March 17             | 639                         | 45                 |
|                                    | Ciano 67  | March 10             | 563                         | 43                 |
| Barley                             | CM 67     | March 4              | 609                         | 49                 |
|                                    | WI2198    | March 6              | 610                         | 54                 |
| Droughted during grain filling (D) |           |                      |                             |                    |
| Wheat                              | Yecora 70 | March 15             | 293                         | 30                 |
|                                    | Ciano 67  | March 9              | 293                         | 32                 |
| Barley                             | CM 67     | March 3              | 363                         | 38                 |
|                                    | WI 2198   | March 5              | 349                         | 40                 |
|                                    |           | s.e.m.               | 18                          | 0.5                |

Plots were sown on 14 December 1974 and received moderate fertilisation ( $75\ kg\ ha^{-1}\ N$  and  $18\ kg\ ha^{-1}\ P$ ). A split plot design (main plot = water regime, sub plot = cultivar) with three replications was used. Each sub plot was  $15\ m \times 1.8\ m$ .

observed  $GDW$ ), as proposed by Gallagher *et al.*<sup>4,5</sup>. But the calculation suggests practically zero contribution in all situations of Table 2. This has been confirmed for irrigated wheat by other experiments with many cultivars at CIANO over several years (ref. 10 and R.A.F., unpublished). The apparent discrepancy when compared to the  $^{14}C$  results of Table 2 probably arises because the absence of net change in the non-grain dry weight after anthesis, does not preclude translocation of pre-anthesis assimilate to the grain, balanced by the accumulation of post-anthesis assimilate in non-grain parts. The fact that the fraction of  $^{14}C$  in the grain from post-anthesis labelling averaged only 0.78 (Fig. 1, Table 2) supports this explanation. Also sampling in other irrigated wheat crops at CIANO showed that from anthesis to maturity the dry weight of the peduncle and chaff increases, while that of the leaves and the remainder of the stem decreases. If these latter decreases are taken to reflect assimilate translocated to the grain, the contribution agrees reasonably well with the 13% value in Table 2.

Our results point to a smaller value of  $P_1$  than that reported by Gallagher *et al.*<sup>4,5</sup>. Notwithstanding the discussion of the preceding paragraph, it is likely that in many situations the method of these authors overestimates  $P_1$ , because the dry

**Table 2** Estimation of the proportion of grain yield derived from pre-anthesis assimilate ( $P_1$ )

| Variable             | Frequently watered |        | Droughted |        |
|----------------------|--------------------|--------|-----------|--------|
|                      | Wheat              | Barley | Wheat     | Barley |
| Pre-anthesis period  |                    |        |           |        |
| Start date*          | 24                 | 15     | 23        | 14     |
| Start $TDW$ †        | 282                | 410    | 282       | 410    |
| Anthesis $TDW$       | 760                | 750    | 610       | 655    |
| $\Delta TDW_1$       | 478                | 340    | 328       | 245    |
| $F_1$                | 0.14               | 0.19   | 0.24      | 0.27   |
| $GDW_1$              | 64                 | 65     | 79        | 67     |
| Post-anthesis period |                    |        |           |        |
| Maturity $TDW$       | 1,358              | 1,367  | 888       | 1,038  |
| $\Delta TDW_2$       | 598                | 617    | 278       | 383    |
| $F_2$                | 0.74               | 0.75   | 0.78      | 0.83   |
| $GDW_2$              | 445                | 465    | 215       | 317    |
| Overall              |                    |        |           |        |
| $GDW_1 + GDW_2$      | 509                | 530    | 294       | 384    |
| $P_1$                | 13%                | 12%    | 27%       | 17%    |
| Observed $GDW$       | 601                | 609    | 293       | 356    |

$TDW$ , total dry weight ( $\text{g m}^{-2}$ );  $GDW$ , grain dry weight ( $\text{g m}^{-2}$ ).  $TDW$  measured by harvests ( $0.30 \text{ m}^2$  quadrat per sub plot) at dates corresponding to those when  $^{14}\text{C}$  labelling was carried out, and by harvest ( $1.80 \text{ m}^2$  per sub plot) at maturity; anthesis  $TDW$  estimated by interpolation, and the increase in total dry weight during the pre-anthesis ( $\Delta TDW_1$ ) and post-anthesis ( $\Delta TDW_2$ ) periods was calculated.  $F_1$  and  $F_2$  are the average proportions of total crop  $^{14}\text{C}$  at maturity in the grain for labelling in the pre- and post-anthesis periods, respectively, and were obtained from Fig. 1. The contribution to grain yield of pre-anthesis ( $GDW_1$ ) and post-anthesis ( $GDW_2$ ) assimilate in absolute terms ( $\text{g m}^{-2}$ ) are equal to  $F_1 \times \Delta TDW_1$ , and  $F_2 \times \Delta TDW_2$ , respectively. The pre-anthesis assimilate contribution as a percentage ( $P_1$ ) is given by  $100 \times GDW_1 / (GDW_1 + GDW_2)$ . The data were averaged for the two cultivars of each species.

\*Time (d) before anthesis of first  $^{14}\text{C}$  labelling date.

†Average of both water regimes used.

weight change in non-grain parts includes tissue, principally leaves, lost due to decay, wind and so on. At other sites in Mexico,  $P_1$  estimated with this method for various non-droughted wheat cultivars ranged from 12 to 98%<sup>11</sup>, and clearly increased as site rainfall and disease incidence increased. We believe this reflects tissue loss, which even at CIANO with no rain or disease amounted to at least  $100 \text{ g m}^{-2}$ . Consequently the results from Nottingham<sup>4,5</sup> may have overestimated the importance of the pre-anthesis contribution of assimilate to grain yield. Even in the most suitable conditions for the expression of this contribution in our experiment (well grown wheat crops, post-anthesis drought reducing kernel weight 30%), it amounted to only 27% of final yield or  $79 \text{ g m}^{-2}$  according to the  $^{14}\text{C}$  data.

This work was carried out with the cooperation of the International Maize and Wheat Improvement Centre, Mexico, and Cornell University.

F. BIDINGER

ICRISAT, I-11-256 Begumpet,  
Hyderabad, 500-016, A.P., India

R. B. MUSGRAVE

Department of Agronomy, Cornell University,  
Ithaca, New York 14853

R. A. FISCHER

CSIRO Division of Plant Industry, P.O. Box 1600,  
Canberra City, A.C.T. 2601, Australia

Received 9 May; accepted 6 October 1977.

1. Archbold, H. K. *Nature* 156, 70–73 (1945).
2. Thorne, G. N. *Rep. Rothamsted exp. Sta.* 1973 2, 5–25 (1974).
3. Lupton, F. G. H. *Ann. appl. Biol.* 64, 363–74 (1969).
4. Gallagher, J. N., Biscoe, P. V. & Scott, R. K. *J. appl. Ecol.* 12, 319–36 (1975).
5. Gallagher, J. N., Biscoe, P. V. & Hunter, B. *Nature* 264, 541–2 (1976).
6. Biscoe, P. V., Gallagher, J. N., Littlejohn, E. J., Monteith, J. L. & Scott, R. K. *J. appl. Ecol.* 12, 295–318 (1975).
7. Fischer, R. A. & Laing, D. R. *J. agric. Sci., Camb.* 87, 113–22 (1976).

8. Birecka, H. & Dakic-Wlodkowska, L. *Acta Soc. Bot. Poloniae* 35, 637–62 (1966).
9. Rawson, H. M. & Evans, L. T. *Aust. J. Agric. Res.* 22, 851–63 (1971).
10. Aguilar, I. & Fischer, R. A. *Agrociencia (Mexico)* 21, 185–98 (1975).
11. Midmore, D. J. thesis, Univ. Reading (1976).

## Physiological energetics of cock-crow

ALTHOUGH much is known about bird song in the behavioural context there is no exact information concerning the efficiency of the communication system itself in terms of energy usage. Little attention has been paid to study of energy relationships in animal communications. The domestic fowl (*Gallus domesticus*) is a convenient subject on which to make physiological measurements of the energy involved in sound production, because of its size and its readiness to make calls in the laboratory, and because of the sheer volume of sound it produces. I have made such measurements and report them here.

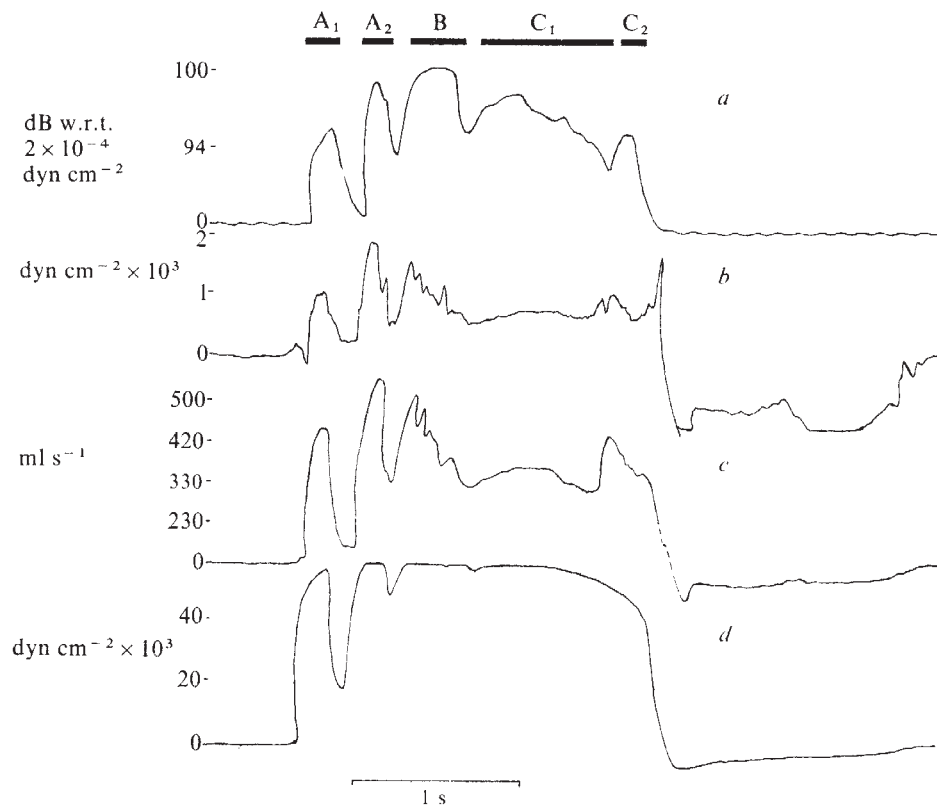
Cock-crow is visibly and audibly one of the most vigorous and spectacular forms of avian vocalisation. The maximum root-mean-square (r.m.s.) sound pressure level attained during the performance is equivalent to a value of 100 dB at a distance of 1 m (Fig. 1a). This corresponds to a peak value of 103 dB. This very impressive achievement is accompanied by enormous rises in air-sac pressure and air flow in the trachea (Fig. 1c, d). Pressure rises to  $60 \text{ cm H}_2\text{O} = 6 \times 10^4 \text{ dyn cm}^{-2}$  and averages  $55 \text{ cm H}_2\text{O} = 5.5 \times 10^4 \text{ dyn cm}^{-2}$ . Air flow reaches  $500 \text{ ml s}^{-1}$  and averages  $\sim 350 \text{ ml s}^{-1}$  during the cycle.

Compared with a normal expiration, these values represent increases of  $\sim 100$ -fold and 15-fold for pressure and flow respectively. This indicates that, concomitant with sound production, there is a rise in airway resistance of some 6–7-fold. This rise is attributed to a change in the configuration of the syrinx, the sound-producing organ, involving, amongst other things, a narrowing of the syringeal lumen by the intruding tympaniform membranes<sup>1–3</sup>. In keeping with this idea, Fig. 1b indicates that the pressure changes in the trachea are comparatively small and, therefore, that virtually the whole of the pressure head generated in the air sacs is spent during the passage of air through the syrinx.

The fluid energy losses in the syrinx are accounted for mainly by frictional dissipation, but also to a lesser extent by conversion of a part of the energy into the mechanical vibration of the external tympaniform membranes, and hence into airborne vibrations. It is possible to estimate approximately the efficiency of this conversion process, and therefore of the effectiveness of the lung air sac–syrinx apparatus as an audio-generator, by comparing the total fluid energy losses incurred by the effort with the amount of sound energy finally produced. The work performed in expelling air from the lung air sac system to the atmosphere is estimated by measuring graphically the area enclosed by the pressure–volume curve during a single crowing cycle (Fig. 2d). The volume curve is itself obtained by graphical integration of the linearised flow curve. The final figure is  $\sim 3.5 \text{ W s}$ . The duration of the crow is 2 s, so the average work rate is 1.75 W. This does not, however, represent total crowing effort since no account has been taken of the internal work of the expiratory muscles lost as contraction heat, and the external work done in moving the viscera enclosing the air sacs.

To calculate the sound energy radiated, the sound pressure level curve given in Fig. 1a is first re-drawn as an equivalent power curve ( $0 \text{ dB} = 2 \times 10^{-4} \text{ dyn cm}^{-2} = 10^{-10} \text{ W cm}^{-2}$ ) which is then integrated over the crowing cycle to yield the total energy per  $\text{cm}^2$  (Fig. 3). The resultant value, assuming a maximum r.m.s. sound pressure level of 100 dB, works out at  $\sim 9 \times 10^{-7} \text{ W s}$ . This is equivalent to an average sound reading throughout the cycle of 95 dB. The total area

**Fig. 1a**, Average sound pressure level, at a distance of 1 m, during crow. Recordings were made with a Brüel and Kjaer 4161 omnidirectional microphone connected to a Nagra 1V SJ recorder and levels were confirmed by a Brüel and Kjaer 2209 precision sound level meter. Recordings were passed through a Grass 7P3 fast integrator to give a display of the sound pressure waveform. Bars above indicate division of crowing cycle into sections: A<sub>1</sub>, A<sub>2</sub>, introductory; B, transitional; C<sub>1</sub>, plateau; C<sub>2</sub>, final kick. **b**, Pressure changes in trachea. **c**, Air-flow rate. The flow meter was implanted into the trachea 2 cm caudal to the larynx. The calibration is nonlinear. **d**, Pressure changes in interclavicular air sac. Flow and pressure measurements were made with Grass PT5 manometers. Scale bar, 1 s.



over which the sound is radiated is calculated approximately by assuming, as is done in the case of human speech<sup>4</sup>, that the sound isobars are arranged in front of the mouth as concentric hemispheres. At a distance of 1 m, the total area of such a hemisphere is  $\sim 6 \times 10^4 \text{ cm}^2$ ; therefore, the total energy radiating from the mouth is  $\sim 9 \times 10^{-7} \times 6 \times 10^4 \text{ W s} = \sim 0.054 \text{ W s}$ . The average sound power is therefore  $0.054/2 = 0.027 \text{ W}$  or 27 mW. This figure is approximately 27 times the maximum speech power radiated during very loud human conversation<sup>5</sup>.

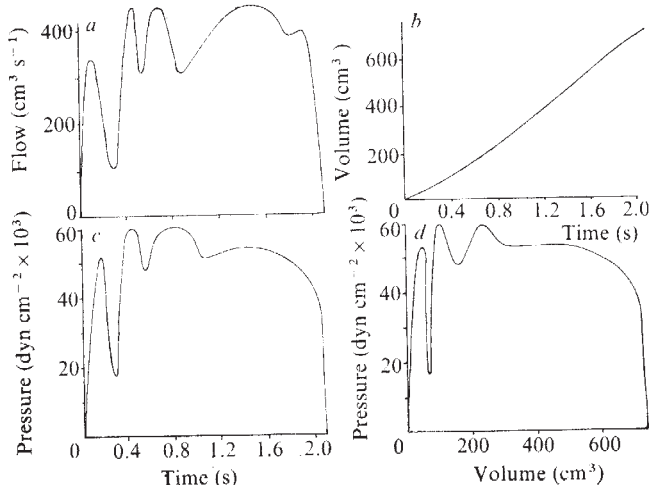
The efficiency of the sound-producing process, measured as the ratio of sound to fluid power, is  $0.055/3.5 \times 100 = 1.6\%$ . The overall efficiency, given in terms of the ratio of sound output to total physical exertion measured by the energy consumption of the driving muscles, is less than

1.6% since it also includes on the input side thermodynamic energy wastage in the contracting tissue and energy spent moving the viscera.

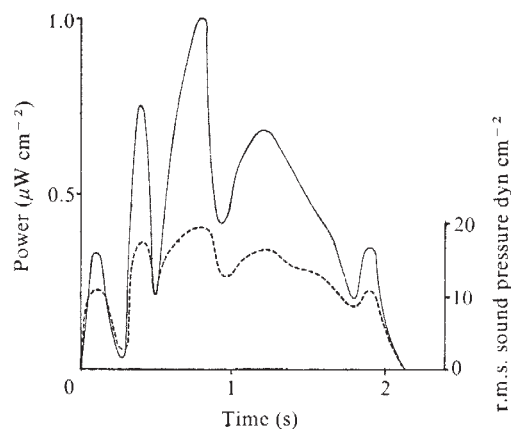
Songbirds are capable of generating sounds which seem to be much louder in proportion to their body size, and therefore in proportion to their power resources, than chickens (unpublished observations). The efficiency of the sound-producing process in these smaller birds therefore seems to be greater. This may be related to well-known differences in the structure of the syrinx<sup>6</sup>, particularly the assemblage of intrinsic muscles which the chicken syrinx altogether lacks, and which seems to be responsible for accurately regulating both the shape of the syrinx and the flow of air through it.

I know of no other estimates of the efficiency of sound production of birds. It is, however, interesting to compare the figures for the chicken with those given for certain insects which use an entirely different mechanism of sound production, based on the rubbing of one part of the body against another. Crickets of the family Gryllinae can pro-

**Fig. 2** Calculation of efficiency of sound production process. **a**, Linearised flow wave form. **b**, Volume curve obtained by integration of (a). **c**, Air-sac pressure waveform. **d**, Air-sac pressure-volume curve obtained from (c) and (b). The area enclosed by the curve gives the fluid work done during a single crowing cycle, given by  $\int PdV$ —work.



**Fig. 3** Power curve (—), obtained from sound pressure curve (---). The total energy per  $\text{cm}^2$  is obtained by measuring the area beneath the power curve.





duce  $6 \times 10^{-2}$  mW of sound with an overall efficiency of 5% whilst the mole cricket *Gryllotalpa vineae* produces 1.2 mW with a remarkable efficiency of 30% (ref. 7). These insects are clearly more economical in their use of energy than the fowl in which the overall efficiency is probably no greater than 1%. Nothing definite can be said as to the position of songbirds on the efficiency scale, although it seems to be greater than 1%.

J. H. BRACKENBURY\*

University of Cambridge,  
Sub-Department of Veterinary Anatomy,  
Tennis Court Road,  
Cambridge, UK

Received 18 July; accepted 6 October 1977.

\*Present address: Department of Biology, University of Salford, Salford, UK.

1. Brackenbury, J. H. *J. Zool., Lond.* (in the press).
2. Gaunt, A. S., Stein, R. C. & Gaunt, S. L. *J. exp. Zool.* **183**, 241–262 (1973).
3. Gaunt, A. S., Gaunt, S. L. L. & Hector, D. H. *Condor* **78**, 208–223 (1976).
4. Sacia, C. F. *Bell Syst. Tech. J.* **4**, 627 (1925).
5. Dunn, H. K. & Farnsworth, D. W. *J. acoust. Soc. Am.* **10**, 184 (1939).
6. Warner, R. W. *J. Zool., Lond.* **168**, 381–393 (1972).
7. Bennet-Clark, H. C. *Nature* **234**, 255–259 (1971).

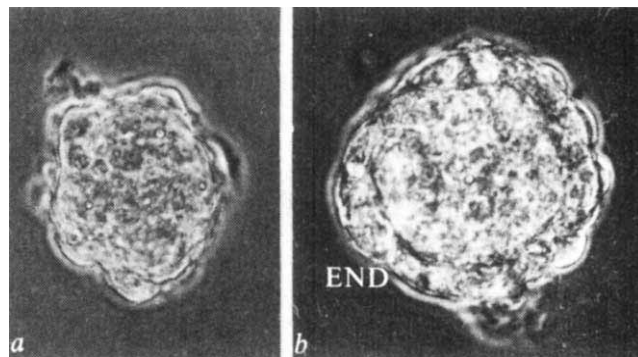
## Regeneration of endoderm by ectoderm isolated from mouse blastocysts

CELL determination can be described as the progressive restriction in developmental options during embryogenesis<sup>1</sup>. Mouse embryo cells, in particular, remain undetermined for a relatively long period, and the fate of cells seems to be labile throughout the period of cleavage<sup>2–4</sup>. Micro-surgical analysis of mouse blastocysts, however, has shown that the developmental potential of inner cell mass and trophoblast cells becomes restricted during blastocyst formation, so that neither of these early cell types retains the option of differentiating into the other<sup>5–7</sup>. We have now investigated whether the primary ectoderm, which forms in the inner cell mass on day 5 of gestation<sup>8,9</sup>, retains the option of differentiating into endoderm. We found that when ectoderm was isolated from mouse blastocysts by immunosurgery<sup>10</sup> it could regenerate an outer layer of endoderm during further culture. The capacity of isolated ectoderm to differentiate into endoderm depended both on the initial mass of the embryo and on the time that the ectoderm was isolated from the inner cell mass.

Viable ectoderm cells can be isolated from pre-implantation mouse embryos by immunosurgery of blastocysts and thin inner cell masses<sup>11</sup>. Ectoderm thus isolated and cultured on feeder layers for 1–22 d differentiates into mesodermal and ectodermal derivatives, although it does not reform visceral endoderm<sup>12</sup>. But because the development of the inner cell mass in intact blastocysts requires a critical mass of tissue<sup>13</sup>, we investigated the developmental potential of ectoderm with increased mass isolated from aggregation chimaeras.

The survival of ectoderms isolated from chimaeric embryos and incubated in our culture conditions depended strongly on the number of embryos used to form the chimaera (Table 1). Ectoderms isolated from single embryos and chimaeras composed of two embryos degenerated; most of the ectoderms from chimaeras composed of four embryos survived, although only a few of these regenerated an endoderm layer. Most ectoderms isolated from chimaeras composed of eight or 10 embryos, however, survived and regenerated an outer layer of endoderm cells (Table 1, Fig. 1a, b). These cells had the characteristic apical vacuoles and microvilli of visceral endoderm cells (Fig. 2a, b) of mouse embryos *in vivo*<sup>8,17</sup> and *in vitro*<sup>9,18</sup>.

Although isolated ectoderms seemed by phase contrast microscopy to be essentially free of contaminating endoderm cells, it was possible that some endoderm cells



**Fig. 1** Regeneration of endoderm by ectoderm isolated from chimaeric mouse blastocysts. *a*, Ectoderm just after immunosurgery of isolated inner cell mass (24 h after late blastocyst stage) (phase contrast,  $\times 310$ ). *b*, Ectoderm cultured for 48 h after immunosurgery. END, endoderm layer. (Phase contrast,  $\times 310$ .)

survived immunosurgery and rapidly proliferated to form a new endodermal layer. To determine whether regenerated endoderm arose from contaminating endoderm cells, we labelled cells after the second immunosurgery with <sup>3</sup>H-thymidine, and compared the grain counts per nucleus in ectoderm and regenerated endoderm after 48 h of culture (Table 2). The dilution of <sup>3</sup>H-thymidine labelling indicates a turnover time of about 24 h for the DNA of these cells. Assuming that there is little re-utilisation of salvaged DNA<sup>19</sup>, these results indicate a cell-cycle time for both ectoderm and endoderm cells *in vitro* of approximately 24 h. Therefore there would have to be a substantial number of contaminating endoderm cells (about 24) to produce the mean number of endoderm cells (96) present at 48 h (Table 2). We serially sectioned nine ectoderms fixed just after immunosurgery and found an average of three (0,0,1,1,2,2,3,8 and 10) contaminating endoderm cells for each ectoderm. Even these few endoderm cells may not have survived, because isolated ectoderms commonly lost a few dead outer cells during the 24 h following immunosurgery. We conclude that there were too few viable contaminating endoderm cells to account for the final number of regenerated endoderm cells and that both the ectoderm and the endoderm present at the end of the experiment arose from the ectoderm cell population.

The time at which ectoderm was isolated from the

**Table 1** Effect of number of embryos in chimaeric blastocysts on endoderm regeneration by isolated mouse ectoderm cells

| No. of embryos in chimaera | Total no. | No. of degenerated ectoderms | Alive, no endoderm | Endoderm + ectoderm |
|----------------------------|-----------|------------------------------|--------------------|---------------------|
| 2                          | 14        | 14                           | 0                  | 0 (0%)              |
| 4                          | 18        | 5                            | 11                 | 2 (11%)             |
| 8                          | 19        | 1                            | 7                  | 11 (57.9%)          |
| 10                         | 31        | 2                            | 6                  | 23 (74.2%)          |

Two-cell embryos were obtained from mated Dub:(ICR) mice (Flow Laboratories) after superovulation with pregnant mares' serum gonadotropin and human chorionic gonadotropin. They were cultured in modified standard egg culture medium as before<sup>13</sup>. At the four- to eight-cell stage, between two and 10 embryos were aggregated to form chimaeras in the presence of phytohaemagglutinin<sup>14</sup>. At the late blastocyst stage (96-h culture from the two-cell stage) inner cell masses were isolated by immunosurgery<sup>10</sup>, embryos were incubated in rabbit anti-mouse L-cell (NCTC Clone 929) serum (1:1 dilution for 25 min) and guinea pig complement (Gibco) (1:5 dilution for 45 min). Ectoderms were isolated by further immunosurgery after inner cell masses had been cultured for another 24 h in modified Eagle's basal medium<sup>13,16</sup> containing 5% foetal calf serum and 5% newborn calf serum. Ectoderms were cultured individually for 48–72 h in drops of medium, with or without paraffin oil, in bacteriological Petri dishes (Falcon) to prevent attachment and outgrowth of ectodermal cells on the substrate. They were scored for the presence of an endodermal layer under the dissecting microscope ( $\times 140$ ).

**Table 2** Grain counts on nuclei of  $^3\text{H}$ -thymidine-labelled cells before and after endoderm regeneration

| Cell type | Hours cultured | No. of samples | Mean no. of cells per sample* | Mean % labelled | Mean no. of grains per nucleus $\pm$ s.e.m. |
|-----------|----------------|----------------|-------------------------------|-----------------|---------------------------------------------|
| Ectoderm  | 0              | 2              | 292                           | 74              | $43.8 \pm 2.4$                              |
|           | 24             | 3              | 192                           | 86              | $17.7 \pm 1.4$                              |
|           | 48             | 2              | 253                           | 67              | $9.2 \pm 1.5$                               |
| Endoderm  | 24             | 1              | 109                           | 61              | $15.9 \pm 1.7$                              |
|           | 48             | 3              | 96                            | 73              | $9.1 \pm 2.6$                               |

Two-cell embryos were cultured overnight and made into chimaeras (10 embryos each) as described in Table 1. Inner cell masses were obtained by immunosurgery from these embryos at the late blastocyst stage; ectoderms were obtained from inner cell masses after 24 h. Ectoderms were labelled with  $^3\text{H}$ -thymidine ( $10^{-2}$   $\mu\text{Ci ml}^{-1}$ , specific activity 45 Ci mmol $^{-1}$ ) for 4 h, washed and cultured for the designated time before hypotonic treatment and fixation in glacial acetic acid and 95% ethanol (1:1) in watch glasses essentially as described<sup>20</sup>. At these concentrations of  $^3\text{H}$ -thymidine there was no effect on cell viability<sup>25</sup>. Autoradiographic exposure was for 3 d. Endoderm and ectoderm layers were separated before fixation by enzyme treatment as before<sup>21</sup>.

\*These numbers may be underestimates because of losses during separation or fixation.

inner cell mass by immunosurgery also strongly influenced its ability to regenerate endoderm. When ectoderms were isolated from inner cell masses 24 or 48 h after isolation of the inner cell masses from late blastocysts the ectoderms regenerated a layer of endoderm. But virtually all ectoderms isolated 72 h after isolation of inner cell masses from late blastocysts failed to regenerate endoderm, and instead they degenerated within 48 h of further culture (Table 3).

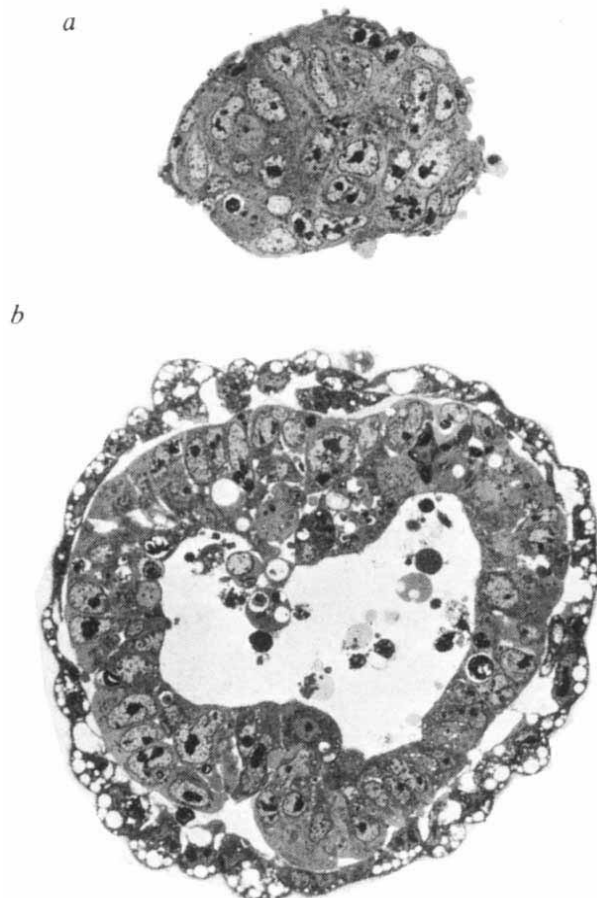
Our results show that the inner cells of the isolated

mouse inner cell mass—the embryonic ectoderm—retain the capacity to differentiate *in vitro* into endoderm for a period well beyond the time that endoderm differentiation normally occurs. Because isolated ectoderms in our experiments regenerate visceral endoderm 48 h after the late blastocyst stage, there seems to be a period of at least 2 d before ectoderm of the mouse inner cell mass undergoes any restriction in its developmental fate. Although failure of ectoderm to regenerate endoderm when isolated after 72 h of culture could imply that ectoderm itself is sensitive to the immunosurgery and culture procedures, we consider this possibility unlikely because of the high incidence of development when it is isolated at earlier stages (Table 3). We conclude that loss of the ability to form visceral endoderm is due to a determinative event between 48 and 72 h after the late blastocyst stage.

By contrast, the developmental potential of visceral endoderm cells seems already to be restricted at the late blastocyst stage. Endoderm cells that are transferred to host blastocysts colonise only the extra-embryonic regions, while ectoderm cells colonise the entire conceptus<sup>8</sup>. Also, visceral endoderm isolated from mouse and rat egg cylinders fails to survive when transferred to ectopic sites<sup>21–23</sup>. The definitive endoderm of the foetus therefore does not seem to arise from the visceral endoderm<sup>6,21–23</sup>.

The major difference between our approach and that of others who did not observe endoderm regeneration by isolated ectoderms<sup>11,12</sup> was our use of chimaeric embryos to increase the mass of the isolated tissues and our prevention of ectoderm attachment to substrate. We interpreted the disintegration of ectoderms isolated from single or paired embryos as an indication that survival of functional inner cells requires a critical mass of tissue which does not exist once trophoblast and endoderm cells are destroyed by immunosurgery. It may be possible, however, to circumvent this requirement in part by

**Fig. 2** Thick section of ectoderm before and after endoderm regeneration. *a*, Ectoderm fixed and sectioned just after immunosurgery (24 h after late blastocyst stage) (bright field,  $\times 430$ ). *b*, Ectoderm cultured for 48 h after immunosurgery. Note complete outer layer of endoderm and central cavity containing cell debris (bright field,  $\times 430$ ). Ectoderms were isolated from chimaeric blastocysts as described in Table 2. Samples were fixed in glutaraldehyde and osmium and embedded and sectioned as before<sup>9</sup>.

**Table 3** Effect of ICM culture time on formation of endoderm by isolated ectoderms

| Hours of culture of isolated ICM | Total no. of ectoderms | No. of degenerated ectoderms | No. of live ectoderms without endoderm | No. of ectoderms with endoderm |
|----------------------------------|------------------------|------------------------------|----------------------------------------|--------------------------------|
| 24                               | 28                     | 0                            | 6                                      | 22                             |
| 48                               | 14                     | 0                            | 0                                      | 14                             |
| 72                               | 31                     | 29                           | 1                                      | 1                              |

Two cell embryos were cultured overnight and chimaeras (10 embryos each) were formed. Inner cell masses (ICM) were isolated by immunosurgery from blastocysts (Table 1) and cultured for various times before isolation of ectoderm by immunosurgery. Growth was scored after 48 or 72 h. Data was pooled from three experiments.



providing a feeder layer of fibroblasts<sup>12</sup>. The failure of ectoderm even from chimaeras of 10 embryos to survive isolation after 72 h of culture implies either that endoderm has an essential, perhaps nutritional function at this time or that after losing the ability to form endoderm, ectoderm cells must attach to survive or to differentiate further. There did not seem to be any obvious morphological change in the ectoderm, however, that correlated with the loss of ability to regenerate endoderm after 72 h in culture. After 24 h of culture isolated inner cell masses have a homogeneous core of ectoderm cells; after 48 and 72 h, the ectoderm has developed into a pseudostratified columnar epithelium surrounding a proamniotic cavity. The cavity, however, increases from approximately 30  $\mu$ m to 100  $\mu$ m in diameter during this time<sup>26</sup> (and unpublished results).

In view of our results with mouse ectoderm isolated 24–72 h after the late blastocyst stage, it seems that the ability to form visceral endoderm, a trait shared by clonal embryonal carcinoma cells<sup>24</sup>, is lost at a time (7–8 equivalent days of gestation) when the embryonic ectoderm can still develop into all three definitive germ layers when it is transferred to ectopic sites<sup>22,23</sup>.

We thank Kitty Wu for assistance, and Gail Martin, Michael Sherman and Davor Solter for comments. This work was supported by the US ERDA.

ROGER A. PEDERSEN

Laboratory of Radiobiology and  
Department of Anatomy,  
University of California,  
San Francisco, California 94143

AKIKO I. SPINDLE  
LYNN M. WILEY\*

Laboratory of Radiobiology,  
University of California, San Francisco

Received 8 August; accepted 3 October 1977.

\*Present address: Department of Anatomy, School of Medicine, University of Virginia, Charlottesville, Virginia 22901.

- Hadorn, E. *Brookhaven Symp. Biol.* **18**, 148–161 (1965).
- Hillman, N., Sherman, M. I. & Graham, C. J. *Embryol. exp. Morph.* **28**, 263–278 (1972).
- Gardner, R. L. & Rossant, J. in *Embryogenesis in Mammals* 5–18 (Elsevier, Amsterdam, 1976).
- Kelly, S. J. in *The Early Development of Mammals* (eds Balls, M. & Wild, A. E.) 97–105 (Cambridge University Press, Cambridge, 1975).
- Gardner, R. L. J. *Embryol. exp. Morph.* **28**, 279–312 (1972).
- Gardner, R. L. & Papaioannou, V. in *The Early Development of Mammals* (eds Balls, M. & Wild, A. E.) 107–132 (Cambridge University Press, Cambridge, 1975).
- Rossant, J. *J. Embryol. exp. Morph.* **33**, 979–990 (1975); **33**, 991–1001 (1975).
- Snell, G. D. & Stevens, L. C. in *Biology of the Laboratory Mouse* 2nd edn (ed. Green, E. L.) 205–245 (McGraw Hill, New York, 1966).
- Wiley, L. & Pedersen, R. A. *J. exp. Zool.* **200**, 389–402 (1977).
- Solter, D. & Knowles, B. *Proc. natn. Acad. Sci. U.S.A.* **72**, 5099–5102 (1975).
- Strickland, S., Reich, E. & Sherman, M. I. *Cell* **9**, 231–240 (1976).
- Hogan, B. & Tilly, R. *Nature* **265**, 626–629 (1977).
- Snow, M. H. L. *J. Embryol. exp. Morph.* **35**, 81–86 (1976).
- Goldstein, L. S., Spindle, A. I. & Pedersen, R. A. *Radiat. Res.* **62**, 276–287 (1975).
- Mintz, B., Gearhart, J. D. & Guymont, A. O. *Devl Biol.* **31**, 195–199 (1973).
- Spindle, A. I. & Pedersen, R. A. *J. exp. Zool.* **186**, 305–318 (1973).
- Solter, D., Damjanov, I. & Skreb, N. *Z. Anat. Entwickl. Gesch.* **132**, 291–298 (1970).
- Solter, D., Biczysko, W., Pienkowski, M. & Koprowski, H. *Anat. Rec.* **180**, 263–280 (1974).
- Cleaver, J. E. *Thymidine Metabolism and Cell Kinetics* 203–207 (North-Holland, Amsterdam, 1967).
- Tarkowski, A. K. *Cytogenetics* **5**, 394–400 (1966).
- Levak-Svajger, B. & Svajger, A. *Experientia* **27**, 683–684 (1971); *J. Embryol. exp. Morph.* **32**, 445–459 (1974).
- Diwan, S. B. & Stevens, L. C. *J. natn. Cancer Inst.* **57**, 937–942 (1976).
- Grobstein, C. J. *exp. Zool.* **119**, 355–379 (1952).
- Martin, G., Wiley, L. M. & Damjanov, I. *Devl Biol.* (in the press).
- Kelly, S. J. & Rossant, J. *J. Embryol. exp. Morph.* **35**, 95–106 (1976).
- Wiley, L. M., Spindle, A. I. & Pedersen, R. A. *Devl Biol.* (in the press).

Table 1 *In vitro* characterisation of melanoma variants

| Cells | Selection | Surviving colonies* | %Agglutination† | % WGA binding‡ |
|-------|-----------|---------------------|-----------------|----------------|
| F-1   | —         | —                   | 97              | 100            |
| Wa-2  | 2nd       | 10                  | 85              | 80             |
| Wa-3  | 3rd       | 50                  | 66              | 69             |
| Wa-4  | 4th       | > 200               | 24              | 62             |

\*Surviving colonies growing out from  $2 \times 10^6$  cells following incubation for 10 d in selective medium ( $100 \mu\text{g ml}^{-1}$  WGA in Eagle's minimal essential medium supplemented with 10% foetal calf serum) and 2 weeks in normal medium without WGA.

†No. of agglutinated cells/no. of total cells  $\times 100\%$  following incubation of  $1 \times 10^6$  cells per ml with  $10 \mu\text{g}$  WGA per ml for 5 min at room temperature.

‡F-1 control cells (100% binding) bound 275.5 c.p.m. per  $\mu\text{g}$  cell protein after 5 min of incubation of  $1 \times 10^6$  cells per ml with  $5 \mu\text{g}$   $^{125}\text{I}$ -WGA per ml at  $4^\circ\text{C}$  (specific activity  $4.8 \times 10^4$  c.p.m. per  $\mu\text{g}$  WGA—supplied by A. Rapin).

have selected variants of melanoma cells for surface differences in using toxic concentrations of lectins in the hope that such variants would exhibit different metastasising capacities. We report here the isolation of variants from metastasising melanoma cells resistant to toxic concentrations of wheat germ agglutinin (WGA). These variants show reduced to a loss of metastasising capacity and altered surface properties.

Mouse B-16 melanoma cells (adapted to grow in tissue culture<sup>6</sup>; supplied by Dr I. J. Fidler, designated as F-1) were maintained in tissue culture for 10 d in selective medium containing WGA at  $100 \mu\text{g ml}^{-1}$ . The WGA-containing medium was then replaced with normal medium and the surviving cells were allowed to grow into discrete colonies in the ensuing 2–3 weeks. All of the resulting six individual colonies were isolated as clones. One of these six clones was further subjected to the same selection procedure successively to produce first Wa-2, then Wa-3 and Wa-4 clones. Table 1 shows that with each additional selection, more cells survived the selective pressure. This increased resistance to the toxic effects of WGA was clearly confirmed in Fig. 1 which compares the growth rates of the parental line F-1 and one of the Wa-4 lines in different concentrations of WGA.

The metastasising capacity was tested by injecting  $5 \times 10^4$  cells intraperitoneally (i.p.) and scoring the resulting tumours in various organs 2–3 weeks later. All animals developed peritoneal tumour masses with ascites and died within 2–4 weeks of injection. Parental cells metastasised in 92% of the animals to mediastinal and mesenteric lymph nodes, and in 58% to liver, kidneys and lungs, as shown in Table 2. In contrast, Wa-2 cells showed a greatly decreased metastasising capacity, while Wa-3 and Wa-4 cells failed to metastasise at all.

Tumorigenicity was tested by injecting  $2 \times 10^6$  cells intravenously (i.v.) and 2–3 weeks later examining the tumour nodules in the lungs, generally the only site where tumours developed after i.v. injection. With successive WGA selections, the metastasising capacity of the variants decreased before tumorigenicity was reduced (Wa-2) while tumorigenicity began to decrease only at the point where no metastases could be detected (Wa-3 and Wa-4) (Table 2). This decreased tumorigenicity was further confirmed by injecting the tumour cells subcutaneously to eliminate the possible complicating factors affecting the survival of circulating tumour cells. Preliminary experiments showing the same pattern in nude mice rule out the possibility that the decreased tumorigenicity was simply due to an alteration in histocompatibility antigens.

Substantial differences in growth rates could influence the results of assays for tumorigenicity and metastasising capacity. But, although *in vivo* growth rates have not yet been assessed, all cell lines showed similar growth rates in tissue culture, with generation times in the range 14–16 h (Fig. 1).

Some of the surface properties of the variants were shown to be different from the parental cells as reflected by a substantial decrease in agglutinability with WGA (Table 1). In addition,

## Non-metastasising variants selected from metastasising melanoma cells

THE cell surface may be involved in the metastatic process of cancer cells<sup>1</sup>. To determine whether membrane properties influence or reflect the metastasising capacity, it is essential to have metastasising and non-metastasising variants of the same tumour. Lectins which bind specifically to surface carbohydrates<sup>2</sup> have been used successfully as selective agents to obtain cells with membrane carbohydrate alterations<sup>3–5</sup>. We



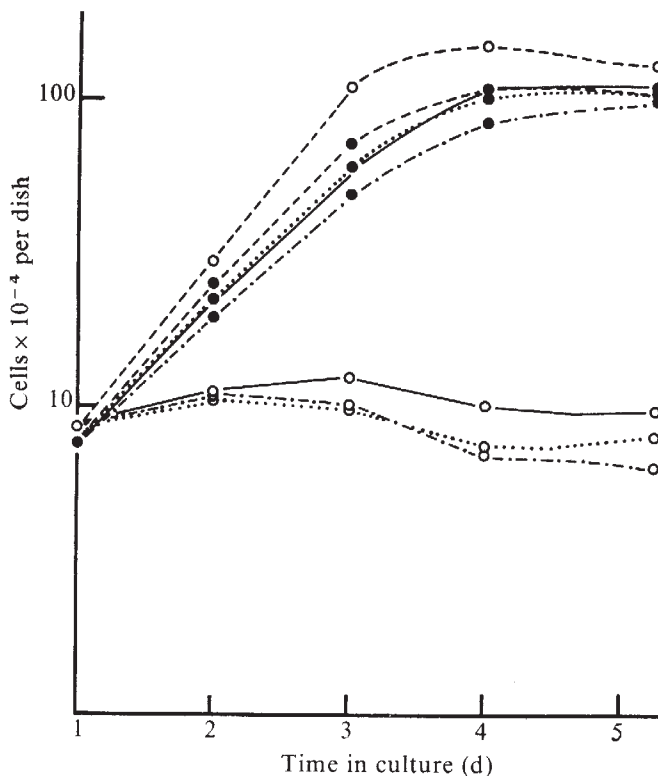


Fig. 1 Growth curves of parental and variant melanoma cell lines in the presence of WGA. Parental F-1 (○) and a Wa-4 variant (●) were grown in 0 (---), 33 (—), 50 (····) and 75 µg WGA per ml (— · — ·) in Eagle's minimal essential medium with 10% foetal calf serum on 10 mm × 35 mm Falcon plastic culture dishes. WGA was added on day 1. Cells were removed daily with trypsin-EDTA and counted in Coulter counter. Each point represents the average value of duplicate dishes.

a decrease in lectin binding was also observed with some lines (Table 1).

The original metastasising line provided by Fidler and the completely non-metastasising variants described here display a maximal difference in metastasising capacity, while they exhibit no significant difference in growth rates *in vitro*. A detailed study of such lines should shed some light on the basic mechanisms underlying the process of metastasis.

By successively passing the B-16 cells through the lungs of the animals, Fidler<sup>8</sup> established lines showing a five- to tenfold increase in metastasising capacity which was defined in his experiment as the number of pulmonary nodules developed after i.v. injection of the tumour cells. Efforts to find surface differences with these lines have been rather disappointing<sup>9</sup>. This could have been because the quantitative difference in metastasising capacity may not have been large enough to allow the detection of biochemical alterations.

Comparison of the properties of metastasising and non-metastasising variants showed metastasising cells to be invasive,

Table 2 Metastasising capacity and tumorigenicity of melanoma variants

| Cells | Metastasising capacity* |              | Tumorigenicity† |
|-------|-------------------------|--------------|-----------------|
|       | Lymph nodes             | Other organs |                 |
| F-1   | 24/26                   | 15/26        | 50/50           |
| Wa-2  | 5/10                    | 0/10         | 15/15           |
| Wa-3  | 0/16                    | 0/16         | 12/14           |
| Wa-4  | 0/16                    | 0/16         | 23/32           |

\*No. of animals showing metastases per total no. of animals tested following i.p. injection of  $5 \times 10^4$  cells. Lymph nodes: mediastinal and mesenteric. Other organs: liver, adrenal glands, kidneys and lungs in this order of frequency.

†No. of animals showing tumours per total no. of animals tested after injection of  $2 \times 10^5$  cells into the tail vein.

where the non-metastasising variants were not invasive. The following preliminary observations were further noted. The non-metastasising variants displayed the greater homotypic adhesion. With successive WGA selection, there seemed to be a gradual increase in the organisation of some of the microstructural elements, from the parental line showing disorganised pattern to Wa-4 lines showing well organised stress fibres. Peptide patterns seen with slab gel electrophoresis following surface labelling were basically similar for the different cell types with the exception of one or two bands which were diminished in the non-metastasising variants.

We thank J. Jenkins and K. Vogel for technical assistance and A. Matter (Hoffmann La Roche) for providing mice. This work was supported by a Swiss National Foundation grant (3.720.76).

TIEN-WEN TAO  
MAX M. BURGER

Department Biochemistry,  
Biocenter, University of Basel,  
CH 4056 Basel, Switzerland

Received 1 September; accepted 30 September 1977.

1. *Fundamental Aspects of Metastasis* (ed. Weiss, L.) (North-Holland, Amsterdam, 1976).
2. Sharon, N. & Lis, H. *Meth. Membr. Biol.* 3, 147–200 (Plenum, New York, 1975).
3. Gottlieb, C., Baenziger, J. & Kornfeld, S. *J. biol. Chem.* 250, 3303–3309 (1975).
4. Stanley, P., Narasimhan, S., Siminovich, L. & Schachter, H. *Proc. natn. Acad. Sci. U.S.A.* 72, 3323–3327 (1975).
5. Meyer, A., Ungkitchanukit, A., Nairn, R. & Hughes, R. C. *Nature* 257, 137–139 (1975).
6. Fidler, I. J. *J. natn. Cancer Inst.* 45, 773–782 (1970).
7. Burger, M. M. *Meth. Enzym.* 32, 615–621 (1974).
8. Fidler, I. J. *Nature* 242, 148–149 (1973).
9. Nicolson, G. L. *et al.* in *Cell and Tissue Interactions* (eds Burger, M. M. & Lash, J.) (Raven, New York, 1977).

## Activation of complement by variant-specific surface antigen of *Trypanosoma brucei*

ANTICOMPLEMENT factors have been shown to occur in certain parasitic infections of man and animals. Hammerberg *et al.*<sup>1</sup> demonstrated the presence of a factor from *Taenia taeniarformis* capable of activating complement through the alternate pathway. Similar findings have been reported for a substance in the cercarial coat of *Schistosoma mansoni*<sup>2</sup>. In trypanosomiasis, no such parasitic activity has been reported although hypocomplementaemia has been recognised as a consistent feature of the disease. Marked reductions in complement C4 and C3 were observed in infections with *Trypanosoma lewisi*<sup>3</sup>, *Tr. rhodesiense*<sup>4</sup> and *Tr. congolense*<sup>5</sup>. One possible mechanism is that the decrease of these complement components results from activation by antigen-antibody complexes. Deposits of immunoglobulins and complement have been demonstrated in kidneys of mice infected with *Trypanosoma brucei*<sup>6,7</sup> and monkeys infected with *Tr. rhodesiense*<sup>4</sup>. We present here evidence for an alternative mechanism operating through direct activation of human complement, via the classical pathway, by an isolated variant-specific surface antigen of *Tr. brucei*.

*Tr. brucei* grown in normal rats was isolated from infected blood using the method of Lanham<sup>8</sup>. In an initial experiment four preparations were used. The first fraction contained live trypanosomes; the second fixed parasites (1% formalin) and the third and fourth were 12,000g (30 min) supernatant and pellet fractions from trypanosomes sonicated on ice for 1 min. All these fractions were tested for the ability to activate human complement in a haemolytic assay<sup>9</sup>. The results showed that each of the four fractions depressed the provided CH<sub>50</sub> (CH<sub>50</sub>, 50% haemolytic dose) by at least 40% (Table 1). Decomplementation also occurred when bovine or rat serum was used as the source of complement. Similar supernatant and pellet fractions derived from *Tr. congolense* were also active (Table 1). Since these preparations may have contained antigen-antibody complexes, analogous fractions were derived from cloned organisms grown in lethally-irradiated mice (900 R) and tested as before. The

**Table 1** Ability of various preparations from *Tr. brucei* and *Tr. congolense* to depress complement activity

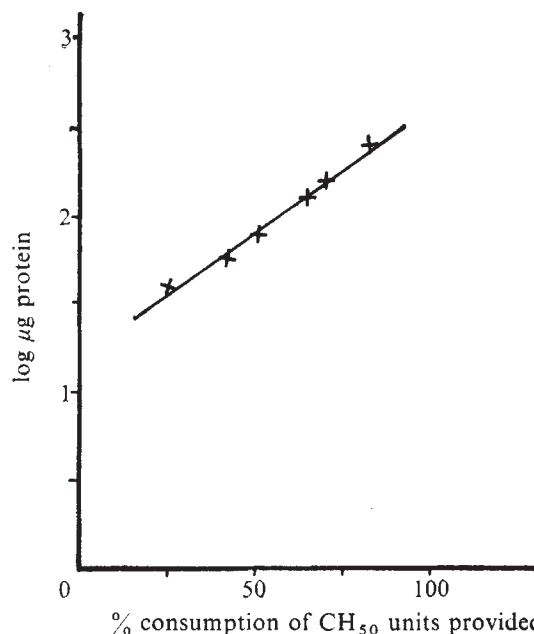
| Parasite              | Preparation | % Consumption of CH <sub>50</sub> per ml |
|-----------------------|-------------|------------------------------------------|
| <i>Tr. brucei</i>     | Live        | 57                                       |
|                       | Formalised  | 53                                       |
|                       | Supernatant | 41                                       |
|                       | Pellet      | 58                                       |
| <i>Tr. congolense</i> | Supernatant | 47                                       |
|                       | Pellet      | 52                                       |

Each preparation used in the assays was derived from  $4 \times 10^7$  parasites and contained in 0.2 ml of veronal-buffered saline. Human serum, obtained from normal donors, was adjusted to provide approximately 10 CH<sub>50</sub> per ml. One ml of the diluted serum was mixed with each fraction or buffer and incubated for 1 h at 37 °C. After incubation the particulate substances were removed by centrifugation and the residual complement titre (CH<sub>50</sub> per ml) determined according to the method of Rapp and Borsos<sup>9</sup>. Sheep cells, sensitised with rabbit haemolysin, were used at a concentration of  $2.5 \times 10^8$  per ml.

results were similar, indicating that the observed utilisation of complement was independent of immune complexes.

The results obtained using the supernatant fraction suggested that the variant antigen, a major soluble glycoprotein, could be involved. To examine this hypothesis a variant specific surface antigen from *Tr. brucei*, clone 052, was prepared using the method of Cross<sup>10</sup>. The parasites were grown from a stabilate in lethally-irradiated mice, followed by passage for 2–3 d in normal rats and then isolated. The purified glycoprotein had a pI of 6.9 and ran as a single band of 64,000 apparent molecular weight in sodium dodecyl sulphate (SDS)–7.5% polyacrylamide gels. These values are in agreement with published data<sup>10</sup>. No host proteins were detected in the preparation by immunodiffusion. This variant antigen, tested in a haemolytic assay as before, caused extensive de complementation using as little as 50 µg protein. A dose–response curve indicated that the de complementation was a logarithmic function of the amount of protein added (Fig. 1). Bearing in mind the extensive differences postulated among *Tr. brucei* surface antigens<sup>10</sup>, a second variant antigen from clone 055 was tested. This glycoprotein was also active.

A further series of experiments was performed to determine the mode of activation of complement. It has been shown by Fine *et al.*<sup>11</sup> that, since the classical and properdin pathways of complement activation have different cation requirements (calcium and magnesium for the former and only magnesium for the latter), EDTA and EGTA may be used to distinguish between the two. EDTA chelates both Ca<sup>2+</sup> and Mg<sup>2+</sup> ions while EGTA binds Ca<sup>2+</sup> effectively but has little affinity for Mg<sup>2+</sup>. In the following experiment, 270 µg of the purified antigen were incubated, for 1 h at 37 °C, with 1 ml of normal human serum (NHS, approximately 320 CH<sub>50</sub> per ml) made 10 mM with either EDTA or EGTA. The residual complement activity was measured as before. The results showed that there was no de complementation in the presence of these chelators, suggesting that activation of complement proceeds by the classical pathway. But, a particulate fraction (1.2 mg protein) prepared from the same clone and tested similarly, depressed complement activity by 72% in the presence of EGTA.

**Fig. 1** A sample of 1 ml of diluted normal human serum (10 CH<sub>50</sub> ml<sup>-1</sup>), was incubated with various amounts of the variant antigen, or buffer, for 1 h at 37 °C. The residual haemolytic titre (CH<sub>50</sub> ml<sup>-1</sup>) was determined using the method of Rapp and Borsos<sup>9</sup>.

This would indicate the presence, in the trypanosome, of another factor capable of activating the alternate pathway.

To confirm the results obtained using specific metal chelators, C1, C4, C2 and C3 activities were measured. The variant antigen was incubated with normal human serum and the residual activities of these individual components estimated using methods described by Rapp and Borsos<sup>9</sup>. The titres were calculated as the number of effective molecules per ml of the individual component being assayed<sup>9</sup>. The results showed preferential utilisation of C1, C4 and C2, with only a moderate change in C3, indicating classical pathway of activation (Table 2).

To test the ability of the variant antigen to cause permeability changes in the skin normal female rats, shaved over the dorsal region were given 0.5 ml of 1% brilliant blue R 10 min before intradermal inoculation of 20 µg of the variant antigen, histamine (5 µg), or buffer. Marked extravasation of dye occurred within 10 min at sites inoculated with variant antigen or histamine. Control sites showed only minimal extravasation of the dye at the site of needle trauma. Administration of an antihistamine (mepyramine maleate 30 mg per kg body weight), before the intradermal inoculations completely abolished the reaction to histamine and to variant antigen. These results suggested that variant antigen mediated production of anaphylotoxins had occurred.

Our observations indicate that complement activation by variant antigen, possibly released during the destruction of trypanosomes at each parasitaemic wave, might be an important contributory factor in depression of complement levels in man<sup>12</sup>

**Table 2** Consumption of some individual complement components in normal human serum incubated with a variant antigen of *Tr. brucei*

| Component assayed | Effective molecules provided | Cellular intermediate | Source of other components | % consumption |
|-------------------|------------------------------|-----------------------|----------------------------|---------------|
| C1                | $7.4 \times 10^{12}$         | EAC4hu*               | C2hu, gp C-EDTA 0.04 M     | 61            |
| C4                | $4.8 \times 10^{12}$         | EAC1gp†               | C2hu, gp C-EDTA 0.04 M     | 95            |
| C2                | $5.0 \times 10^{10}$         | EAC1gp4hu             | gp C-EDTA 0.04 M           | 90            |
| C3                | $1.4 \times 10^{11}$         | EAC1gp4hu             | C2, 5, 6, 7hu<br>C8, 9, gp | 11            |

Normal human serum was diluted with gelatin-veronal buffer containing 0.0015 M Ca<sup>2+</sup> and 0.0005 M Mg<sup>2+</sup> to provide the above effective molecules of individual components. One ml was incubated with 270 µg of the variant antigen, or buffer alone, for 1 h at 37 °C. The residual activities of individual components were then assayed using cellular intermediates and functionally pure components from Cordis Laboratories, Miami. The cellular intermediates were used at a concentration of  $1 \times 10^8$  cells per ml.

\*hu, Human.

†gp, guinea pig.

and in animals suffering from trypanosomiasis. Since recent experiments have shown that complement is essential in antibody-mediated destruction of trypanosomes<sup>13</sup>, the hypocomplementaemic state resulting from the activation would favour the survival of the parasites.

Papamichail *et al.*<sup>14</sup> and others have shown that complement plays a part in antibody production to thymus-dependent antigens. Whether or not trypanosome antigens are thymus-dependent is uncertain. But, such low complement levels are likely to affect the antibody response of the host to trypanosome antigens.

Finally, evidence of increased vascular permeability in the form of perivascular oedema and swelling and degeneration of vessel walls, is a consistent pathological finding in animals infected with *Tr. brucei*<sup>15,16</sup> and *Tr. congolense*<sup>17</sup>. While these changes might result from antigen-antibody reactions, our observation suggesting interaction between variant antigen and complement, producing a vascular permeability factor, should also be considered as a mechanism in the pathogenesis of these and other lesions which occur in African trypanosomiasis<sup>18</sup>.

We thank Venny Shitakha, G. W. Mpimbaza and J. Mayanja-Katende for technical assistance.

*Note added in proof:* Activation of complement unidentified factors from *Tr. congolense* and *Tr. lewis* (Nielsen, K. & Shepard, J. *Experientia* 33, 769-770 (1977)).

A. J. MUSOKE

A. F. BARBET

International Laboratory for Research on Animal Diseases,  
PO Box 30709,  
Nairobi, Kenya

Received 27 June; accepted 7 October 1977.

1. Hammerberg, B., Musoke, A. J., Hustead, S. T. & Williams, J. F. in *Pathophysiology of Parasitic Infections* (ed. Soulsby, E. J. L.) 233-240 (Academic, New York, 1976).
2. Machado, A. J., Gazzinelli, G., Pellegrino, J. & Dias Da Silva, W. *Exp. Parasit.* 38, 20-39 (1975).
3. Jarvinen, J. A. & Dalmasso, A. P. *Infection Immunity* 4, 894-902 (1976).
4. Nagle, R. B. *et al.* *Am. J. Trop. Med. Hyg.* 23, 15-26 (1974).
5. Kobayashi, A. & Tizard, I. R. *Tropenmed. Parasit.* 27, 411-417 (1976).
6. Lambert, P. H. & Houba, V. in *Progr. Immun.* 11, 57-67 (1974).
7. Murray, M., Lambert, P. H. & Morrison, W. I. *Medicine Malad. Infect.* 12, 638-641 (1975).
8. Lanham, S. M. *Nature* 218, 1273-1274 (1968).
9. Rapp, H. J. & Borsos, T. in *Molecular Basis of Complement Action* (Appleton-Century-Crofts, New York, 1970).
10. Cross, G. A. M. *Parasitology* 71, 393-417 (1975).
11. Fine, D. P., Marney, S. R., Colley, D. G., Sergeant, J. S. & Des Prez, R. M. *J. Immunology* 109, 807-809 (1972).
12. Greenwood, B. M. & Whittle, H. C. *Clin. exp. Immunol.* 24, 133-138 (1976).
13. Diggs, C. *et al.* *J. Immunol.* 116, 1005-1009 (1976).
14. Papamichail, M., Guierrez, C., Embling, P., Holbrow, E. J. & Pepys, M. B. *Scand. J. Immunol.* 4, 343-347 (1975).
15. Murray, M., Murray, P. K., Jennings, F. W., Fisher, E. W. & Urquhart, G. M. *Res. Vet. Sci.* 16, 77-84 (1974).
16. Goodwin, L. G. *Trans. R. Soc. Trop. Med. Hyg.* 82-88 (1971).
17. Banks, K. L. *Infect. Immun.* (Submitted).
18. Murray, M. in *Progr. Immunol.* 11, 181-192 (1974).

## Induction of epidermal transglutaminase by hydrocortisone in chick embryonic skin

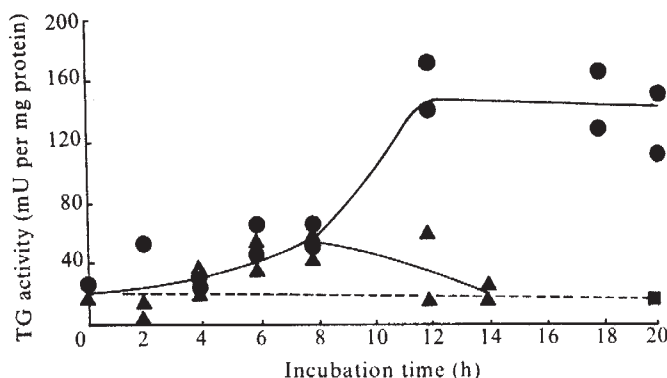
WE demonstrated previously that the epidermal cells of chick embryos are induced to differentiate by glucocorticoids<sup>1,2</sup>. Hydrocortisone ( $0.01 \mu\text{g ml}^{-1}$ ) added to a serum-free chemically-defined medium produced an intense keratinised layer over the uppermost cells<sup>3</sup> and many tonofilaments within the basal and intermediate cells<sup>4</sup> of the epidermis after 4 d in culture of 13-d-old embryonic chick tarsometatarsal skin. Within 12-16 h of hydrocortisone addition it is possible to show the synthesis of a urea-mercaptoethanol soluble, glycine-rich and metabolically stable epidermal structural protein<sup>5</sup> and the accumulation of a urea-mercaptoethanol insoluble epidermal protein(s)<sup>4</sup>. These changes of protein metabolism induced by hydrocortisone paralleled those seen during *in ovo* keratinisation of the epidermis<sup>4,5</sup>. The amino acid composition of the two proteins was very similar suggesting that the soluble protein is converted to the insoluble one in the process of epidermal keratinisation of chick embryonic skin. We show here that the levels of transglutaminase (TG), an enzyme that possibly mediates cross-linking of keratinous polypeptide chains by a  $\epsilon$ -( $\gamma$ -glutamyl) lysine bond<sup>6-8</sup>, increases both during *in ovo*

development of the epidermis and during hydrocortisone-induced *in vitro* keratinisation of chick embryonic cultured skin.

Table 1 shows that TG activity in the undifferentiated epidermis of 13-d-old embryonic chick skin was only 7 mU per mg protein, compared with 53 mU per mg protein in the heavily keratinised epidermis of 19-d-old embryonic chick skin. Furthermore, when 13-d-old embryonic chick skin was cultured for 4 d in a chemically-defined medium with or without hydrocortisone, TG activities of the steroid-treated epidermis were 52-107 mU per mg protein, whereas the activity in control epidermis was only 3-30 mU per mg protein. Figure 1 shows that after culture for 20 h the specific activities of TG in the treated epidermis were seven times those in control tissue. During the cultivation epidermal TG activities started to rise sharply after incubation for 8 h in the presence of the steroid. But, when actinomycin D was added at  $4 \mu\text{g ml}^{-1}$ , a concentration that blocks more than 99% of RNA synthesis, no increase of TG activity was observed, indicating that TG induction by hydrocortisone requires RNA synthesis. A similar effect was seen with cultured chick embryonic neural retina<sup>9</sup>; hydrocortisone added to the medium induced glutamine synthetase followed by *in vitro* differentiation of the tissue and actinomycin D abolished the enzyme induction. Further analyses are necessary to clarify whether hydrocortisone stimulates the synthesis of TG mRNA in our skin culture; however glucocorticoids were reported to induce the synthesis of specific enzymes through the increased synthesis of their mRNA for cultured rat hepatoma cells<sup>10</sup> and for rat liver<sup>11</sup> while many other kinds of cells, tissues and organs require the steroids for induction or maintenance of their differentiated function<sup>12</sup>.

It should be noted that hydrocortisone induced TG several hours before the induction of soluble epidermal structural protein and this earlier increase of the enzyme activity coincided exactly with that seen during *in ovo* development of the epidermis (data not shown). These findings suggest that the enzyme may be involved in a series of finely organised processes of epidermal cell differentiation. But more detailed studies are required to test the hypothesis that soluble structural proteins are cross-linked or polymerised by epidermal TG to insoluble high molecular ones composing tonofilaments to culminate in keratinisation of epidermal cells. One of the problems under study is whether

**Fig. 1** Changes of the specific activities of epidermal transglutaminase (TG) of 13-d-old chick embryonic skin during the cultures in the presence or absence of inducer and suppressor. Methods of skin cultivation and TG determination were essentially as described in Table 1. One of the pair-mate explants after preculture for 1 d was incubated with hydrocortisone (HC;  $0.01 \mu\text{g ml}^{-1}$ ) and the other with the steroid plus actinomycin D (Boehringer Mannheim;  $4 \mu\text{g ml}^{-1}$ ). ■, Epidermal TG activities of the skin cultured without HC; ●, epidermal TG activities of the skin cultured with HC; ▲, epidermal TG activities of the skin cultured with HC plus actinomycin D. Values of each point are the mean of duplicated determinations for a pool of five sheets of the epidermis.





**Table 1** Increase of epidermal transglutaminase activities (mU per mg protein) with keratinisation *in ovo* and *in vitro*

| <i>In ovo</i>                                    |                                             | <i>In vitro</i>                                   |                                           |
|--------------------------------------------------|---------------------------------------------|---------------------------------------------------|-------------------------------------------|
| Undifferentiated epidermis from 13-d-old embryos | Keratinised epidermis from 19-d-old embryos | Undifferentiated epidermis without hydrocortisone | Keratinised epidermis with hydrocortisone |
|                                                  |                                             | Expt 1                                            | 26<br>30                                  |
| 7 ± 1(5)*                                        | 53 ± 4(4)*                                  | Expt 2                                            | 9<br>3                                    |
|                                                  |                                             |                                                   | 93<br>107<br>52<br>52                     |

Pair-mate explants from right and left tarsometatarsal regions of 13-d-old chick embryos were cultured for 4 d in a chemically-defined medium, BGJb supplemented with ascorbate, by the Millipore filter-roller-tube method<sup>3</sup>. One of the pair-mate explants was cultured in the medium containing hydrocortisone sodium hemisuccinate (Upjohn) 0.01 µg ml<sup>-1</sup>, and the other was cultured in medium without the steroid. The epidermis of cultured explants and that from the same part of chick embryos in the normal course to hatching were each obtained by incubating the skin at 37 °C for 10–40 min in Eagle's minimum essential medium supplemented with 10% calf serum and 900 PU per ml Dispase I (a bacterial neutral protease, Godo Shusei). Crude enzyme preparation from the epidermis was prepared by a modification of the method of Ogawa and Goldsmith<sup>6</sup>. The epidermis was homogenised in 10 mM Tris-acetate (pH 7.5)–1 mM EDTA by the use of a glass homogeniser with five strokes at intermediate speed. This and all further operations were at 0–4 °C. The homogenates were centrifuged for 30 min at 12,000g at 4 °C. The sediments were extracted two additional times in the same manner and the supernatants were combined. The supernatant fluid taken in a cellulose tubing (Visking) was concentrated with sucrose. The concentrated supernatant was dialysed against a buffer of 5 mM Tris-acetate (pH 7.5) with 1 mM EDTA. Transglutaminase activity was assayed for the dialysed supernatant as an enzyme preparation by measuring the incorporation of [1,4-<sup>14</sup>C]putrescine (Radiochemical Centre, Amersham; 60 mCi mmol<sup>-1</sup>) into α-casein (according to Hammarsten; Merck) after the method of Goldsmith and Martin<sup>13</sup>. The assay mixture contained 0.11 M glycine-NaOH (pH 10.0) with 0.56 mM EDTA, 10 mM CaCl<sub>2</sub>, 5 mM dithiothreitol, 9 µM labelled putrescine and 0.15% casein. Aliquots (0.5 ml) of the assay mixture were added to 100 µl of each enzyme preparation. The mixtures were incubated for 30 min at 37 °C. The reaction was stopped by the addition of 0.6 ml of 10% trichloroacetic acid. The precipitate was collected on Whatman GF/C filters and the precipitated radioactivity was measured in toluene containing 0.4% 2,5-diphenyloxazole (Dotite) and 0.01% 1,4-bis(2-(5-phenyloxazolyl))-benzene (Dotite) with a Aloka Liquid Scintillation Spectrometer. An enzyme unit is defined as the amount of enzyme that catalyses the incorporation of 1 nmol of putrescine into casein in 30 min. Specific activity is defined as enzyme units per mg protein. Protein was measured by the technique of Lowry *et al.*<sup>14</sup> with bovine serum albumin as a standard. Values are averages of duplicated determinations for a pool of five sheets of the epidermis.

\*Mean ± standard error. Figures in parentheses represent the number of separate experiments.

there are much more ε-(γ-glutamyl) lysine cross-links in the insoluble protein than in the soluble protein of the chick embryonic epidermis.

This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.

AKIKO OBINATA

Department of Physiological Chemistry,  
Faculty of Pharmaceutical Sciences,  
University of Tokyo, Hongo, Tokyo 113

HIROYOSHI ENDO

Department of Physiological Chemistry,  
Faculty of Pharmaceutical Sciences,  
Teikyo University, Sagamiko, Kanagawa 199-01, Japan

Received 28 July; accepted 10 October 1977.

- Endo, H. *et al.* in *Biochemistry of Cutaneous Epidermal Differentiation* (eds Seiji, M. & Bernstein, A.) 359–378 (University of Tokyo Press, Tokyo, 1977).
- Sengel, P. in *Morphogenesis of Skin* 133–136 (Cambridge University Press, Cambridge, 1976).
- Sugimoto, M. & Endo, H. *J. Embryol. exp. Morph.* **25**, 365–376 (1971).
- Sugimoto, M., Tajima, K., Kojima, A. & Endo, H. *Dev. Biol.* **39**, 259–307 (1974).
- Kojima, A., Sugimoto, M. & Endo, H. *Dev. Biol.* **48**, 173–183 (1976).
- Ogawa, H. & Goldsmith, L. A. *J. biol. Chem.* **251**, 7281–7288 (1976).
- Buxman, M. M. & Wuepper, K. D. *J. invest. Dermatol.* **65**, 107–112 (1975).
- Abernethy, J. L., Hill, R. L. & Goldsmith, L. A. *J. biol. Chem.* **252**, 1837–1839 (1977).
- Moscona, A. A., Moscona, M. H. & Saenz, N. *Proc. natn. Acad. Sci. U.S.A.* **61**, 160–167 (1968).
- Steinberg, R. A., Levinson, B. B. & Tomkins, G. M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2007–2011 (1975).
- Schutz, G., Killewich, L., Chen, G. & Feigelson, P. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1017–1020 (1975).
- Sugimoto, M., Kojima, A. & Endo, H. *Dev. Growth Different.* **18**, 319–327 (1976).
- Goldsmith, L. A. & Martin, C. M. *J. invest. Dermatol.* **64**, 316–321 (1975).
- Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265–275 (1951).

## Ionophore-mediated calcium influx effects on the post-synaptic muscle fibre membrane

THE physiological chemotransmitter acetylcholine and its structural analogues such as carbamylcholine are capable of producing a sustained blockade of neuromuscular transmission when applied to the neuromuscular junction for prolonged periods. This neuromuscular blockade is not due to a depolarisation of the postsynaptic muscle fibre membrane but rather is generally ascribed to the inactivation, or 'desensitisation' of cholinergic receptor molecules in the postsynaptic membrane<sup>1,2</sup>. Various kinetic models involving agonist, and receptor

molecules postulated to exist in different conformational states, have been proposed to account for the time course of the desensitisation process<sup>2,3</sup>. These models do not incorporate the acceleration of desensitisation by calcium ions<sup>4,5</sup> and they do not predict correctly the observed effects of certain cholinergic antagonists on desensitisation. Consequently, alternative models for desensitisation have been proposed. In one of them calcium ions accumulate at the interior surface of the post-synaptic membrane and bind to the cholinergic receptors, thereby causing desensitisation by producing or sustaining some inactive receptor conformation<sup>6</sup>. We tested the model for desensitisation which involves internal calcium binding. This was accomplished by facilitating calcium ion flux across the post-synaptic membrane using the divalent cation ionophore A23187, while simultaneously producing a rapid desensitisation with iontophoretic application of carbamylcholine. Our results provide further evidence that calcium ions are a significant factor in the molecular mechanism of desensitisation.

A23187 is a carboxylic ionophore with high affinity for divalent cations and little or no affinity for univalent cations<sup>7</sup>. It is capable of transporting calcium ions across otherwise highly impermeable artificial and biological membranes by an essentially electroneutral process<sup>8</sup>. As a probe in research on contractile systems, A23187 has been used to release stored calcium from isolated sarcoplasmic reticulum vesicles<sup>9,10</sup>. A23187 has also been reported to increase the twitch tension of barnacle muscle fibres without altering the electrical properties of the fibre membranes<sup>11</sup>.

Using established single fibre stimulation and recording techniques, we investigated the effects of a wide range of A23187 concentrations on the electrical behaviour of frog (*Rana pipiens*) cutaneous pectoris or sartorius muscles employed in our desensitisation experiments. Resting and action potentials were recorded at junctional and extra-junctional sites and in addition, miniature end-plate potentials (m.e.p.s) were recorded at the junctions.

Because A23187 is only slightly soluble in aqueous media such as frog Ringer solution, a stock solution was prepared in 100% ethanol and appropriate quantities were added to the Ringer solution. Control experiments showed that ethanol, at the highest concentrations involved (0.5%), had an insignificant effect on the resting potentials, action potentials and m.e.p. frequencies recorded from muscle fibres. A small increase in

**Table 1** Effects of bath-applied A23187 on miniature end-plate potential frequency

| Expt | A23187 Concentration ( $\mu$ M) | Exposure time (min) | M.e.p. frequency (Hz) | No. of fibres |
|------|---------------------------------|---------------------|-----------------------|---------------|
| 1    | 0 (control)<br>0.28             | Continuous          | $0.30 \pm 0.22^*$     | 10            |
|      |                                 |                     | $0.57 \pm 0.13$       | 12            |
| 2    | 0 (control)<br>0.56             | Continuous          | $1.1 \pm 0.7$         | 6             |
|      |                                 |                     | $1.4 \pm 1.3$         | 6             |
| 3    | 0 (control)<br>3.0              | 10                  | $2.0 \pm 0.3$         | 10            |
|      |                                 |                     | $4.8 \pm 1.2$         | 11            |
| 4    | 0 (control)<br>7.5              | 5                   | $1.3 \pm 0.6$         | 10            |
|      |                                 |                     | $6.0 \pm 3.9$         | 14            |

Control solution was HEPES buffered Ringer solution (composition: 112.4 mM Na<sup>+</sup>, 2.5 mM K<sup>+</sup>, 1.8 mM Ca<sup>2+</sup>, 117.1 mM Cl<sup>-</sup>, 3.0 mM HEPES buffer; pH, 7.4). A23187 solutions contained 0.02, 0.04, 0.2 and 0.5% ethanol respectively.

\*Mean  $\pm$  s.d.

m.e.p. frequency was obtained when the Ringer solution contained 1% ethanol; this agrees with published data on the effects of alcohol on neuromuscular systems<sup>12,13</sup>. In our experiments involving the production of desensitisation the alcohol concentration never exceeded 0.04%. Control experiments showed that at this concentration ethanol had no effect on desensitisation.

Two different methods were used to apply the ionophore to the muscle fibres in both electrical behaviour studies and desensitisation studies. In the first method A23187 stock solution was added to HEPES-buffered Ringer solution (see Table 1 for composition) forming a suspension of the ionophore. After making recordings from muscle fibres bathed in ionophore-free Ringer solution, the low concentration A23187 solutions were continuously bath-applied during which time a second group of recordings was obtained. With solutions containing high concentrations of A23187 (3–15  $\mu$ M) application to the muscle was limited to a few minutes following which the bath was changed to ionophore free Ringer solution and thereafter additional recordings were made. This procedure was used to avoid production of 'retraction clots' and other mechanical disturbances presumably caused by an excessive ionophore-promoted increase in Ca<sup>2+</sup> influx which could lead to activation of the contractile system.

Many fibres exposed to A23187 for 5–10 min at concentrations above 8  $\mu$ M contracted irreversibly (formed retraction clots) and others twitched spontaneously for several minutes after the ionophore was removed from the bath. Retraction clots were also produced in fibres perfused with 6  $\mu$ M to 8  $\mu$ M A23187 solutions.

In the second method of applying the A23187, a quantity of the ionophore stock solution was added to calcium-free Ringer solution. A small region of a muscle immersed in ordinary Ringer solution was then microperfused with the Ca-free A23187 solution for 1–2 min. Recordings were made at the perfusion sites after allowing the ordinary Ringer solution to diffuse back into the ionophore treated region. This perfusion method was used in order to avoid mechanical artefacts (see above).

From bath application experiments we found that A23187, at concentrations below 1.6  $\mu$ M, had no effect on the muscle fibre resting and action potentials. With cells loaded at higher ionophore concentrations (up to 15  $\mu$ M the resting potentials were slightly diminished and the action potentials showed reduced rates of rise and fall. Statistical analysis, however, indicated that these differences are of doubtful significance, in contrast with the marked effects observed in experiments utilising the divalent cation ionophore X537A (ref. 14).

All of the A23187 solutions used in our experiments caused some increase in m.e.p. frequency; typical values are given in Table 1. The small increase from 1.1 to 1.4 s<sup>-1</sup> at 0.56  $\mu$ M A23187 is not statistically significant, but the differences do become

significant at the higher ionophore concentrations. Often, with A23187 applied at high concentrations (bath or perfusion), prolonged high frequency bursts (showers) of m.e.ps were observed, followed by decreased frequencies below control levels.

The above findings indicated that A23187 functioned as an electroneutral calcium ionophore in the skeletal muscle preparations, since calcium ion influx triggers both quantal release of acetylcholine from nerve terminals<sup>15</sup> and activates the contractile mechanism in muscle fibres<sup>16</sup>. Further support for this conclusion was obtained when perfusion of 6  $\mu$ M A23187 solution (Ca<sup>2+</sup>-free) failed to produce either retraction clots or increased m.e.p. frequencies in fibres bathed in Ringer solution in which the Ca<sup>2+</sup> had been replaced with Mg<sup>2+</sup>.

Desensitisation experiments were performed using constant current iontophoretic pulses of carbamylcholine<sup>17</sup>. The applied current pulses flowing through the iontophoretic micropipette and the corresponding evoked membrane potential changes were monitored continuously. Endplate regions of high sensitivity to the agonist were first located using single 5–10 ms duration iontophoretic pulses giving approximately 1 mV depolarisation responses, thus avoiding any preliminary desensitisation effects<sup>18</sup>. Trains of iontophoretic pulses ranging from 0.4 s duration at 1.0 Hz to 1.0 s at 0.5 Hz were then generated.

During delivery of the first few iontophoretic pulses, the evoked membrane potential changes produced by the applied carbamylcholine generally increased in amplitude over the

**Table 2** Desensitisation half-times,  $\tau_{1/2}$ , for frog skeletal muscle exposed to A23187

| Fibre        | Resting potential (mV) | $\Delta V_{max}$ (mV)§ | $\Delta V_r$ (mV) | $\tau_{1/2}$ (s) |
|--------------|------------------------|------------------------|-------------------|------------------|
| 1, Control*  | 88                     | 14.5                   | 8.8               | 11.5             |
| 1, Perfused† | 83                     | 15.0                   | 9.2               | 8.0              |
| 2, Control   | 85                     | 9.5                    | 5.6               | 16.0             |
| 2, Perfused† | 87                     | 10.6                   | 2.0               | 12.0             |
| 3, Control   | 89                     | 0.6                    | 0.4               | 10.4             |
| 3, Perfused† | 85                     | 0.6                    | 0.4               | 4.2              |
| 4, Control   | 84                     | 2.9                    | 2.9               | —¶               |
| 4, Perfused† | 88                     | 3.5                    | 2.5               | 15.0             |
| 5, Control   | 92                     | 6.3                    | 3.5               | 12.0             |
| 5, Perfused† | 92                     | 4.2                    | 1.0               | 7.0              |
| 6, Control   | 85                     | 23.8                   | 12.6              | 13.7             |
| 6, Bath‡     | 90                     | 20.0                   | 13.7              | 8.2              |
| 7, Control   | 85                     | 36.5                   | 32.5              | 12.7             |
| 7, Bath‡     | 87                     | 16.0                   | 10.6              | 7.9              |
| 8, Control   | 87                     | 9.0                    | 6.0               | 19.0             |
| 8, Bath‡     | 88                     | 10.0                   | 4.0               | 10.5             |
| 9, Control   | 83                     | 7.5                    | 6.0               | 25.0             |
| 9, Bath‡     | 86                     | 11.2                   | 5.0               | 15.0             |
| 10, Control  | 83                     | 10.0                   | 4.8               | 12.0             |
| 10, Bath‡    | 85                     | 11.2                   | 8.5               | 8.5              |
| 11, Control  | 85                     | 11.2                   | 8.0               | 11.5             |
| 11, Bath‡    | 88                     | 12.0                   | 6.2               | 7.5              |
| Means        |                        |                        |                   |                  |
| Control      | 86                     | 12.0                   | 8.3               | 14.4             |
| Experimental | 87                     | 10.4                   | 5.7               | 8.9              |

\*Ringer solution contains 1.8 mM Ca<sup>2+</sup> all fibres below bathed in Ringer solution containing 3.6 mM Ca<sup>2+</sup>.

†After the control period, the fibre was perfused with 2.4  $\mu$ M A23187 solution (Ca<sup>2+</sup>-free ethanol-0.16%) for 1 to 2 min; recordings were then repeated 10 to 15 min after perfusion.

‡After the control period, 0.56  $\mu$ M A23187 (0.04% ethanol) was bath applied to the fibre and recordings were repeated in the presence of the ionophore after waiting 30 to 60 min.

§Maximum depolarisation pulse amplitude (see Fig. 1).

||Final (steady-state) depolarisation pulse amplitude (see Fig. 1).

¶No desensitisation.



first few pulses to a maximum,  $\Delta V_{\max}$  then declined in amplitude to a final steady response  $\Delta V_f$  as illustrated in Fig. 1. The phenomenon of increasing  $\Delta V$  amplitudes which occurs with the first few iontophoretic pulses is not an artefact but represents a physiological potentiation phenomenon<sup>19,20</sup>. The decline in evoked potential amplitude which, occurred after the maximum response was obtained, indicated that desensitisation was in progress. The time  $\tau_i$  for these depolarisations to decline from  $\Delta V_{\max}$  to  $\frac{1}{2}(\Delta V_{\max} + \Delta V_f)$  was taken as a measure of the desensitisation rate. Comparisons of control and ionophore treated fibres were made only between recordings taken at the same postjunctional location on a given fibre because under control conditions there has been found to be a large variation in desensitisation rates amongst endplates from the same muscle<sup>18</sup>.

The results of these desensitisation experiments, presented in Table 2, demonstrated that very low concentrations of A23187 accelerated the desensitisation rate by 30% to 60%. For example, the desensitisation half time for fibre 1, which was 11.5 s in the control, decreased to 8.0 s after perfusion with 2.4  $\mu\text{M}$  A23187 solution. Apparently A23187 did not alter the sensitivity of the fibres to carbamylcholine, because following its application inconsequential random changes in the value of  $\Delta V_{\max}$  occurred. Direct measurement of endplate sensitivities, determined as membrane response in  $\text{mV/nC}^{-1}$  of applied iontophoretic drug, confirmed this conclusion. Thus A23187 did not promote desensitisation by direct antagonistic interaction with the cholinergic receptors.

From the data we and others have obtained showing that A23187 carries calcium ions across excitable membranes, it could be concluded that the substantial acceleration of desensitisation by the ionophore was due to its action in increasing calcium ion influx both in the resting and the carbamylcholine activated postsynaptic membrane. We wondered what increase in intracellular calcium concentration would be produced by the action of A23187 alone. The concentrations of A23187 used in the desensitisation experiments were well below those at which retraction clots or changes in the electrical characteristics of the excitable membrane were produced. Thus if the minimal concentration of  $\text{Ca}^{2+}$  required to activate the contractile mechanism<sup>21</sup> is  $10^{-6}$  M then one might conclude that the  $[\text{Ca}^{2+}]_i$  in these ionophore loaded fibres was less than  $10^{-6}$  M before development of desensitisation by application of carbamylcholine.

Although the ionophore produced increase in  $[\text{Ca}^{2+}]_i$  seems too be small, its effect become evident during application of

desensitising iontophoretic pulses of carbamylcholine which further increase  $\text{Ca}^{2+}$  influx across the postjunctional membrane. We attempted to calculate the increase in  $[\text{Ca}^{2+}]_i$  resulting from the action of A23187 alone but the value obtained was clearly too high and we attributed this apparently erroneous result to the many questionable assumptions involved in the calculation.

In the above speculations it is tacitly assumed that the influx of  $\text{Ca}^{2+}$  ions are free to distribute themselves in the adjacent intracellular fluid. As suggested earlier<sup>6</sup>, however, the possibility exists that during production of receptor desensitisation influxing  $\text{Ca}^{2+}$  ions become bound to anionic sites on the interior surface of the postjunctional membrane and also accumulate in the intracellular fluid immediately adjacent to this membrane. In this model a relatively small amount of influxing  $\text{Ca}^{2+}$  could produce a substantial localised increase in calcium concentration. Such a localised increase in calcium would not necessarily cause a retraction clot because, as can be seen in electron micrographs, the intracellular region immediately adjacent to the postjunctional membrane does not contain contractile elements and in addition, some of the intracellular calcium would not be free.

The effect of the increase in  $[\text{Ca}^{2+}]_i$  produced by A23187 becomes evident during application of desensitising iontophoretic pulses of carbamylcholine to the postjunctional membrane which further increase  $\text{Ca}^{2+}$  influx. We suppose that  $\text{Ca}^{2+}$  ions exert some positive cooperative action in stabilising an inactive receptor state, or a closed ion channel conformation. Binding studies with purified acetylcholine receptor (AChR) have shown that ACh receptor has a fairly high affinity for  $\text{Ca}^{2+}$  and that  $\text{Ca}^{2+}$  influences the affinity of ACh receptor for cholinergic agonists<sup>22-24</sup>. If we assume that the agonist-receptor complex goes spontaneously through a series of active and then inactive states, a small localised increase in  $\text{Ca}^{2+}$  might maintain an ever increasing number of receptors in a complex but inactive state.

This work was supported by grants from the National Institute for Neurological and Communicative Disorders and Stroke (5 R01-04988, 1 F22 NS01438). We thank Dr Robert Hamill (Lilly Research Labs) for supplying A23187, and C. H. Wolfson and J. T. Alexander for technical assistance and helpful discussions.

DAVID I. DEVORE\*  
WILLIAM L. NASTUK

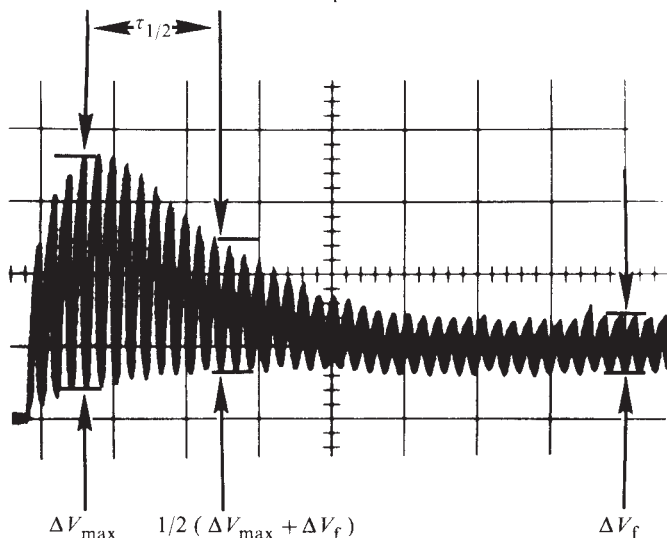
College of Physicians and Surgeons,  
Columbia University,  
New York, New York 10032

Received 7 October 1976; accepted 26 September 1977.

\*Present address: Department of Chemistry, University of California, Berkeley, California 94720.

- Thesleff, S. *Acta physiol. scand.* **34**, 218-231 (1955).
- Katz, B. & Thesleff, S. *J. Physiol., Lond.* **138**, 63-80 (1957).
- Rang, H. P. & Ritter, J. M. *Molec. Pharmac.* **6**, 357-382 (1970).
- Manthey, A. A. *J. gen. Physiol.* **49**, 963-975 (1966).
- Magazanik, L. G. & Vyskocil, F. *J. Physiol., Lond.* **210**, 507-518 (1970).
- Nastuk, W. L. & Parsons, R. L. *J. gen. Physiol.* **56**, 218-249 (1970).
- Reed, P. W. & Lardy, H. A. in *The Role of Membranes in Metabolic Regulation* (eds Mehlman, M. A. & Hanson, R. W.) 111-131 (Academic Press, New York, 1972).
- McLaughlin, S. & Eisenberg, M. A. *Rev. Biophys. Bioengng* **4**, 335-366 (1975).
- Caswell, A. H. & Pressman, B. C. *Biochem. biophys. Res. Commun.* **49**, 292-298 (1972).
- Scarpa, A., Baldassare, J. & Inesi, G. *J. gen. Physiol.* **60**, 735-749 (1972).
- Hainut, K. & Desmedt, J. E. *Nature* **252**, 407-408 (1974).
- Inoue, F. & Frank, G. B. *Br. J. Pharmac. Chemother.* **30**, 186-193 (1967).
- Okada, K. *Jap. J. Physiol.* **17**, 245-261 (1967).
- Devore, D. I. & Nastuk, W. L. *Nature* **253**, 644-646 (1975).
- Miledi, R. *Proc. R. Soc. B183*, 421-425 (1973).
- Huxley, A. F. *J. Physiol., Lond.* **243**, 1-43 (1974).
- Nastuk, W. L. *Fedn Proc.* **12**, 102 (1953).
- Wolfson, C. H. & Nastuk, W. L. *Fedn Proc.* **34**, 404 (1975).
- Hartzell, H. C., Kuffler, S. W. & Yoshikami, D. *J. Physiol., Lond.* **251**, 427-463 (1975).
- Wolfson, C. H. thesis, Columbia Univ. (in preparation).
- Ebashi, S., Endo, M. & Ohtsuki, I. *Q. Rev. Biophys.* **2**, 351-384 (1969).
- Martinez-Carrion, M. & Raftery, M. A. *Biochem. biophys. Res. Commun.* **55**, 1156-1164 (1973).
- Cohen, J. B., Weber, M. & Changeux, J.-P. *Molec. Pharmac.* **10**, 904-932 (1974).
- Chang, H. W. & Neumann, E. *Proc. natn. Acad. Sci. U.S.A.*, **73**, 3364-3368 (1976).

Fig. 1 Typical membrane depolarisations,  $\Delta V$ , produced during iontophoretic application of a 1 Hz train of 0.7 s constant current carbamylcholine pulses to the muscle postjunctional membrane. Note the early facilitation followed by later developing desensitisation. Calibrations: ordinate 0.5 mV per division, abscissa 5 s per division.





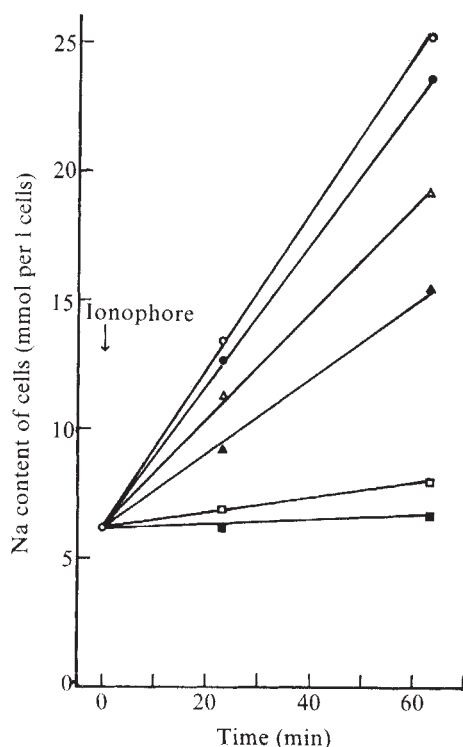
## Does ionophore A23187 mediate Na transport in the absence of divalent cations?

THE divalent cation-selective ionophore A23187 is known to form complexes with monovalent cations<sup>1</sup>, but it is not clear whether such complexes actually translocate across biological membranes. In the absence of Ca and Mg, the ionophore can transport K across the mitochondrial membrane in a manner reminiscent of the nigericin class of ionophores<sup>2</sup>. From spectrofluorometric and bulk organic phase partition data<sup>3</sup>, Pfeiffer and Lardy concluded that Na binds to the ionophore better than K ( $\text{Li} \gg \text{Na} > \text{K} \approx \text{O}$ ) and it is surprising that no observations of this Na form of the ionophore crossing biological membranes have been reported. We have used a recently developed method<sup>4</sup> designed to study the effect of intracellular Mg on the various ion fluxes mediated by the Na pump, and we report here that intact human red cells incubated in buffered isotonic NaCl media containing EDTA, no added divalent cations and the ionophore

**Table 1** Effect of A23187 on unidirectional fluxes of  $^{24}\text{Na}$  in the presence or absence of Mg in low-Na media

| Medium  | Ionophore A23187 | $[\text{Mg}^{2+}]_i$ (M)  | Na fluxes (mmol per l cells per h) |                   |
|---------|------------------|---------------------------|------------------------------------|-------------------|
|         |                  |                           | Influx                             | Efflux            |
| K       | —                | $\sim 3.2 \times 10^{-4}$ | $0.210 \pm 0.014$                  | $0.836 \pm 0.009$ |
| Choline | —                | $\sim 3.2 \times 10^{-4}$ | $0.137 \pm 0.008$                  | $1.377 \pm 0.009$ |
| K       | +                | $3.2 \times 10^{-4}$      | $0.198 \pm 0.008$                  | $0.747 \pm 0.013$ |
| Choline | +                | $3.2 \times 10^{-4}$      | $0.147 \pm 0.006$                  | $1.216 \pm 0.017$ |
| K       | +                | $\sim 10^{-7}$            | $0.290 \pm 0.006$                  | $1.330 \pm 0.047$ |
| Choline | +                | $\sim 10^{-7}$            | $0.292 \pm 0.012$                  | $1.495 \pm 0.021$ |

Fresh, washed human red cells were incubated with or without  $^{24}\text{Na}$  for 4 h at 37 °C in a Tris-buffered medium containing NaCl 75 mM and KCl 75 mM. The cells were then washed and suspended at a haematocrit of about 10% in a medium containing NaCl 5 mM, Tris-Cl 10 mM, Tris-EGTA 0.01 mM, Tris-EDTA 2 mM, ouabain  $10^{-4}$  M and 145 mM of either KCl or choline chloride.  $\text{MgCl}_2$  2.1 mM was present in groups three and four of the table giving the final concentrations of internal  $\text{Mg}^{2+}$  as indicated.  $^{24}\text{Na}$  was added to the suspension of tracer-free cells to measure Na influx. Na fluxes were calculated by multiplying the internal (expressed per l of cell water for efflux) and external (for influx) Na concentrations by the respective rate constants of tracer movement (the influx rate constant was corrected for backflux).



**Fig. 1** Effect of the ionophore A23187 on the uptake of sodium by intact red cells at various Mg concentrations. Red cells from 2-d-old bank blood were washed in Tris-buffered isotonic saline and resuspended at a haematocrit of 10% in a medium initially containing NaCl 150 mM, Tris-HCl 10 mM (pH 7.7 at 37 °C), Tris-EGTA 0.01 mM, Tris-EDTA 2 mM and ouabain  $10^{-4}$  M. The suspension was divided into six equal parts and  $\text{MgCl}_2$  was added from a concentrated stock solution (0.8 M) to give the initial total concentrations reported here ( $[\text{Mg}]_0 = 0$ ). Values of  $[\text{Mg}]_0 = 0$  were (mM):  $\circ$ , 0;  $\bullet$ , 0.5;  $\triangle$ , 1.0;  $\blacktriangle$ , 1.5;  $\square$ , 2.1;  $\blacksquare$ , Control (no ionophore). At  $t = 0$ , 5  $\mu\text{l}$  of absolute ethanol containing 1 mg  $\text{ml}^{-1}$  of ionophore A23187 were added (arrow) per ml of suspension during vigorous magnetic stirring (at 37 °C). Samples of 1 ml were taken at the indicated times. The cells were centrifuged and washed three times with 10 volumes of ice-cold Tris-buffered choline chloride solution. Na and K were determined by flame photometry. Concentrations are expressed in mmol per l original cell volume. The initial concentration of K in the cells was  $89.8 \pm 1.4$  and the final concentration was between 87.5 and 88.6 in the controls and high-Mg conditions and between 81.6 and 84.5 at the lowest Mg concentrations.

A23187, gain Na and swell. This effect was not accompanied by any change in the K permeability and could be prevented or reversed by adding Mg or Ca to the medium. Thus, in certain conditions A23187 may indeed mediate Na transport.

The effect of the ionophore A23187 on the bulk and tracer ( $^{24}\text{Na}$ ) fluxes of Na was investigated in intact red cells suspended in 'Ca-free' media containing ouabain, in order to avoid both Ca-induced changes in K permeability<sup>4,5</sup> and variations in Na-pump rate that occur at different internal Mg concentrations (to be reported elsewhere; see also ref. 6). Figure 1 shows the effect of the ionophore on the Na content of cells suspended in isotonic Na media containing increasing Mg concentrations. The nominally Mg-free cells ( $[\text{Mg}^{2+}]_i \approx 10^{-7}$  M) gained about 18 mmol of Na per l of cells in 1 h, a value about 10 times the Na gain in ionophore-free cells or in cells containing ionophore and normal Mg. The net Na influx is plotted as a function of  $[\text{Mg}^{2+}]_i$  in Fig. 2, showing half-maximum inhibition in the micromolar range. The increase in Na flux was fully reversed by physiological concentrations of internal Mg. Ca was also found to inhibit the Na influx in the presence of the ionophore but this effect was masked to some extent by a separate action of Ca on the Na influx. Ca tended to increase the Na influx compared with Ca-free controls even in the presence of high concentrations of Mg, and made it difficult to assess the precise extent to which Ca inhibited the ionophore-mediated influx.

In order to avoid large changes in ion composition and cell volume, the effect of the ionophore on the unidirectional ( $^{24}\text{Na}$ ) tracer fluxes was investigated in media containing 5 mM Na with either K or choline as the main extracellular cation. The results of one such experiment are reported in Table 1. Here again it can be seen that the ionophore increases the unidirectional Na fluxes only in the absence of Mg. At the same time, however, the K content of the ionophore-treated cells was the same as that of untreated controls in a similar medium. Na efflux is increased relatively more when K is the main external cation, while the effect on the influx of Na is relatively larger in the choline medium. This seems to result from an apparent effect of external K on the mechanism which is responsible for the asymmetry between Na efflux and Na influx in the presence of ouabain<sup>7</sup> and which had been variously attributed in the past to a second Na-pump<sup>8</sup>, to counter-transport of Na

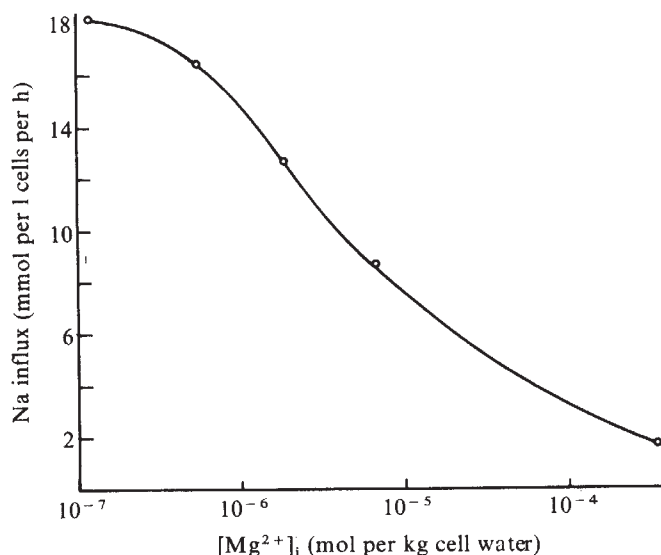


Fig. 2 Net gain of sodium as a function of the intracellular concentration of free Mg in intact red cells in the presence of the ionophore A23187. The Na uptake of Fig. 1 is plotted as a function of the  $Mg^{2+}$  content of the cells calculated from the ionophore-induced equilibrium distribution of Mg as reported previously<sup>3</sup>.

with choline or  $Mg^{7,9}$  or to co-transport of Na with  $K^{10,11}$ . The present results rule out symmetric carrier-mediated transport involving choline, K or Mg, as Na efflux is larger than Na influx in all conditions, whether in the presence or absence of choline, K, or Mg gradients. The persistence of the asymmetric Na fluxes even in 'Mg-free' cells also argues against ATP-fuelled transport associated with a Mg-dependent ATPase<sup>12</sup>.

In order to discriminate between a possible indirect

effect of the ionophore and ionophore-mediated Na flux, we measured net Na influx in cells that had been treated with the ionophore in the presence and absence of Mg and from which the ionophore had been washed away<sup>5,13</sup>. This was done to see whether 'Mg-free' cells with little or no ionophore left could still exhibit the increased Na permeability. The results (Table 2) showed that the 'Mg-free' cells did not have an increased Na permeability unless the ionophore was restored during the final incubation. All the available evidence suggests either that the ionophore can act on the membrane to produce a reversible and selective increase in Na permeability at low  $Mg^{2+}$  or  $Ca^{2+}$  concentrations, an interesting possibility which cannot be ruled out, or, more likely, that the ionophore itself mediates Na transport when the concentration of competing divalent cations is very low.

We thank the Wellcome Trust Foundation and the MRC for providing funds to V.L.L., the Lilly Research Centre for a gift of the ionophore A23187, A. M. Brown for helpful comments and Mrs Gray for technical assistance. P.F. is an MRC Scholar.

PETER FLATMAN  
VIRGILIO L. LEW

Physiological Laboratory,  
University of Cambridge,  
Cambridge, UK

Received 12 August; accepted 3 October 1977.

- Pfeiffer, D. R., Reed, P. W. & Lardy, H. A. *Biochemistry* 13, 4007-4014 (1974).
- Pfeiffer, D. R. & Lardy, H. A. *Biochemistry* 15, 935-943 (1976).
- Flatman, P. & Lew, V. L. *Nature* 267, 360-362 (1977).
- Gardos, G. *Biochim. biophys. Acta* 30, 653-654 (1958).
- Lew, V. L. & Ferreira, H. G. *Nature* 263, 336-338 (1971).
- Bodemann, H. H. & Hoffman, J. F. *J. gen. Physiol.* 67, 547-561 (1976).
- Beaugé, L. *Biochim. biophys. Acta* 401, 95-108 (1975).
- Hoffman, J. F. & Kregenow, F. M. *Ann. N.Y. Acad. Sci.* 137, 566 (1966).
- Beaugé, L. & Ortiz, O. *J. Membr. Biol.* 13, 165-185 (1973).
- Sachs, J. R. *J. gen. Physiol.* 57, 259-282 (1971).
- Lew, V. L. & Beaugé, L. in *Transport Across Biological Membranes* (eds Giebisch, G., Tosteson, D. C. & Ussing, H. H.) vol. 2 (Springer, in the press).
- Proverbio, F. M., Condrescu-Guidi & Whitembury, G. *Biochim. biophys. Acta* 394, 281-292 (1975).
- Sarkadi, B., Szasz, L. & Gardos, G. *J. Membr. Biol.* 26, 357-370 (1976).

Table 2 Effect of ionophore removal on Na influx in the presence and absence of internal Mg

| Pretreatment of the cells                    | Ionophore A23187 during final incubation | Presence of Mg during final incubation | Na influx (mmol per l cells per h) |
|----------------------------------------------|------------------------------------------|----------------------------------------|------------------------------------|
| Controls                                     | —                                        | —                                      | $2.76 \pm 0.104$                   |
|                                              | —                                        | +                                      | $1.60 \pm 0.154$                   |
|                                              | +                                        | —                                      | $25.79 \pm 0.461$                  |
|                                              | +                                        | +                                      | $3.48 \pm 0.259$                   |
| Treated with ionophore in the absence of Mg  | —                                        | —                                      | $3.43 \pm 0.459$                   |
|                                              | —                                        | +                                      | $3.92 \pm 0.382$                   |
|                                              | +                                        | —                                      | $26.99 \pm 0.590$                  |
|                                              | +                                        | +                                      | $3.80 \pm 0.446$                   |
| Treated with ionophore in the presence of Mg | —                                        | —                                      | $1.97 \pm 0.236$                   |
|                                              | —                                        | +                                      | $2.13 \pm 0.346$                   |
|                                              | +                                        | —                                      | $28.84 \pm 0.708$                  |
|                                              | +                                        | +                                      | $2.90 \pm 0.273$                   |

Cells were pretreated with A23187 in the presence and absence of 2.1 mM  $MgCl_2$  in a Tris-buffered isotonic choline chloride medium containing 5 mM NaCl. A low sodium medium was used to prevent changes in the intracellular Na concentration during the preincubation. 5  $\mu$ l of a 1 mg ml<sup>-1</sup> solution of the ionophore in absolute ethanol was added for each ml of the suspension. The haematocrit was about 10% and the pre-incubation was performed at 37 °C for 30 min. The ionophore was absent from the control group, otherwise conditions were identical. The cells were subsequently washed five times in a similar ice-cold solution and resuspended in a medium containing NaCl 150 mM; Tris-HCl 10 mM (pH 7.7 at 37 °C), Tris-EGTA 0.01 mM, Tris-EDTA 2 mM and ouabain 10<sup>-4</sup> M.  $MgCl_2$ , when present, was at 2.1 mM. <sup>23</sup>Na was added to the final incubation medium in order to measure Na influx. When the ionophore was absent, the 'Mg-free' cells did not gain Mg in the Mg-containing media.

## Lowering of hypertension by central saralasin in the absence of plasma renin

A ROLE for the renin-angiotensin system in the development or maintenance of increased blood pressure of spontaneously hypertensive (SH) rats has not been substantiated. These animals are considered the best model for human essential hypertension. Although there have been reports of unusually high concentrations of renin in the plasma of young SH rats<sup>2</sup>, their pathophysiological relevance has been questioned<sup>3</sup> and, as we report here, high blood pressure is maintained even after nephrectomy. But angiotensin from the brain isorenin-angiotensin system (Iso-RAS) could be involved<sup>4,6</sup>. All the components of that system have been identified in brain tissue, including the enzyme iso-renin<sup>4-6</sup>, the substrate, angiotensinogen<sup>7</sup>, angiotensin I and converting enzyme<sup>6,8</sup> which hydrolyses angiotensin I to form the biologically active octapeptide angiotensin II. Angiotensin II is also present in brain tissue<sup>6</sup>. When it is injected intraventricularly into the brain angiotensin II produces a prolonged increase in blood pressure<sup>9</sup>. Therefore, if angiotensin II concentrations are increased in the brains of SH rats they could be involved in the hypertensive state independently of angiotensin levels from renal origin in plasma. Using the competitive antagonist Sar-1-Ala-8-angiotensin II (saralasin acetate, P113, Norwich Pharmacal) in SH and nephrectomized SH rats, we have confirmed this possibility.

Twenty-one SH stroke-prone (SH-SP) rats weighing 220-330 g and 10 age matched Wistar Kyoto control rats (WKY) were anaesthetised by ether, and a 28-gauge stainless steel cannula was implanted stereotactically into the right lateral ventricle. Rats were allowed 3-10 d for recovery and were

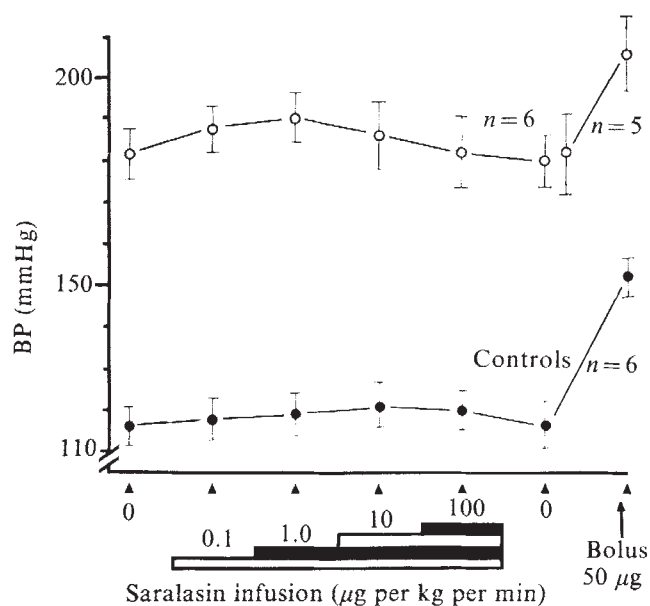
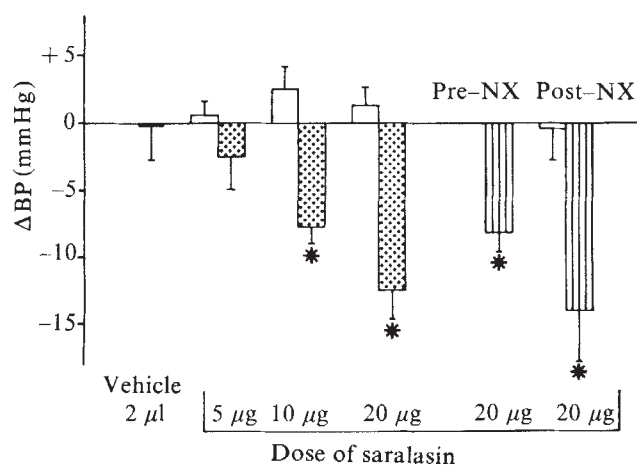
again anaesthetised with ether. A femoral vein catheter (polyethylene No. 10 tubing, Clay Adams, USA) and a femoral artery catheter (polyethylene No. 10 connected to No. 50 tubing) were inserted, ligatured and brought under the skin to exit through the scruff of the neck. A 1-ml blood sample was taken from the arterial catheter for plasma renin assay and simultaneously replaced by 1 ml isotonic saline through the venous catheter. Rats were left awake, freely moving in a wooden box during testing and recording. Blood pressure was recorded by a Statham transducer P 23 b and Hellige pre-amplifiers on to a Brush 220 polygraph. Testing began 90–120 min after blood pressure had become stable.

Ten SH-SP and six WKY male and female rats, all 4-months-old, were tested with intravenous and intraventricular doses of saralasin. Eleven SH-SP rats and four WKY rats (all females, 8-months-old) were nephrectomised and tested. Bilateral nephrectomy was carried out under ether anaesthesia and rats were tested 15–20 h later. Blood samples for plasma renin assays were also taken 15 h after nephrectomy.

To test the potency of the preparation each rat was tested initially with 100 ng angiotensin II intravenously and intraventricularly. After 1 h, saralasin was infused intravenously in cumulative doses of 0.1, 1.0, 10.0 and 100  $\mu\text{g}$  per kg per min for 15 min at each dose. Sixty min after the infusion test a bolus injection of 50  $\mu\text{g}$  saralasin was given intravenously. After 2–3 h recovery when blood pressure was stable, testing began with intraventricular injections of 5, 10 and 20  $\mu\text{g}$  saralasin in 0.5, 1.0 and 2.0- $\mu\text{l}$  volumes, respectively. The order of testing was randomised and there was full recovery of the original blood pressure between tests. Saralasin was dissolved in artificial cerebrospinal fluid and control injections of 2  $\mu\text{l}$  of this vehicle were also given.

The results are illustrated in Figs 1 and 2. The mean arterial blood pressure of SH-SP rats was  $181.3 \pm 9.2$  mmHg. Intraventricular saralasin lowered blood pressure in a dose-response relationship in every rat tested ( $n = 14$ ) (Fig. 1). In the 4-month-old rats 5  $\mu\text{g}$  saralasin produced a decrease of  $2.5 \pm 2.3$  mmHg. The 10- $\mu\text{g}$  dose lowered blood pressure by  $7.8 \pm 1.2$  mmHg and the 20- $\mu\text{g}$  dose by  $12.5 \pm 1.9$  mmHg ( $P < 0.01$  compared with controls by the  $t$ -test). A 20- $\mu\text{g}$  dose of saralasin lowered blood pressure in the 8-month-old rats by  $8.3 \pm 1.4$  mmHg ( $P < 0.01$  compared with controls). The decrease began in each case within 1 min of injection and reached a maximum by

**Fig. 1** Changes in mean arterial blood pressure ( $\Delta$  BP in mmHg) after injections of the angiotensin blocker saralasin into the cerebral ventricles of spontaneously hypertensive rats of the stroke-prone strain and into the cerebral ventricles of normotensive Wistar Kyoto control rats. Doses of intraventricular saralasin are indicated below. Pre-NX: rats tested before and in the same animals 16 h after bilateral nephrectomy (Post-NX). Dotted columns, 4-month-old ( $X \pm \text{s.e.m.}$ ); open columns, age-matched controls; striped columns, 8-month-old.



**Fig. 2** The effect of intravenous saralasin infusions in cumulative doses on the mean arterial blood pressure (BP) of spontaneously hypertensive rats of the stroke-prone strain.

5–7 min. There was recovery back to the original blood pressure within 15–30 min. WKY rats matched for age showed no response to intraventricular saralasin (Fig. 1). When saralasin was given intravenously the two smallest doses caused increase of blood pressure in SH ( $n = 6$ ) and control rats ( $n = 6$ ) ( $P < 0.05$ ) (Fig. 2). Bolus injection of 50  $\mu\text{g}$  saralasin consistently produced an even more marked increase in blood pressure in both SH ( $23.4 \pm 2.5$  mmHg) and control rats ( $35.3 \pm 2.6$  mmHg) (Fig. 2).

To investigate whether circulating plasma angiotensin II of renal origin was involved in the central effect of saralasin we repeated the test in nephrectomised rats. In nephrectomised SH-SP rats the mean lowering of blood pressure by 20  $\mu\text{g}$  saralasin given intraventricularly was  $14.1 \pm 3.8$  mmHg. Intraventricular saralasin had no effect on blood pressure in nephrectomised control rats. Plasma renin activity in these rats was  $0.5 \pm 0.1$  ng angiotensin I  $\text{ml}^{-1} \text{h}^{-1}$ , which is close to detection limits of the assay and lower as compared with the control levels in SH rats of  $5.7 \pm 0.8$  ng angiotensin I  $\text{ml}^{-1} \text{h}^{-1}$ . Nephrectomy did not affect the blood pressure of control or experimental rats. The mean arterial pressure before nephrectomy was  $178.3 \pm 6.2$  mmHg and after nephrectomy it was  $178.0 \pm 6.3$  mmHg.

The response to angiotensin II in SH rats seemed to be greater than in WKY rats, as recently confirmed<sup>15</sup>.

These data implicate angiotensin in the central blood pressure regulation of SH rats. Although the effect was comparatively small, the consistency with which intraventricular saralasin lowered the blood pressure may reflect further factors which tend to maintain blood pressure in these rats. The effect is not readily obtained when blood pressure is lower<sup>16</sup>, nor was it obtained in normotensive control rats. The results also demonstrate that in the absence of peripheral plasma renin after nephrectomy, the blood pressure of SH rats is still increased.

The lowering of blood pressure by central saralasin in the absence of peripheral renin-angiotensin is evidence for a role of the brain iso-renin-angiotensin system in the maintenance of high blood pressure in SH-SP rats. Since its discovery<sup>1,5</sup>, several lines of evidence have suggested that the system has a physiological function. Electrophysiological evidence and recent biochemical binding studies have demonstrated angiotensin receptors in the brain<sup>10–13</sup>. Saralasin slowed the rate of firing of cells responsive to angiotensin<sup>10</sup>. This has been



interpreted to mean that the spontaneous rate was normally maintained at high rates by endogenous angiotensin. A reduction of blood pressure by central infusions or injections of angiotensin antagonists has been reported in other strains of SH rats (New Zealand and Okamoto strains)<sup>14-18</sup>, but negative findings have also been claimed<sup>19</sup>. Without nephrectomy these results were only circumstantial evidence for an action of extrarenal angiotensin. With the data reported here we conclude that extrarenal angiotensin acts on the brain receptors and contributes to the high blood pressure in SH-SP rats. In view of the growing support for the presence of locally formed brain angiotensin<sup>20</sup>, the brain iso-renin system is a likely extrarenal source.

We thank Dr A. Castellion, Norwich Pharmacal and Dr Völger, Röhm Pharma for the gift of saralasin. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 90) and Humboldt Fellowship to M.I.P.

M. IAN PHILLIPS\*  
JOHANNES F. E. MANN  
HIDEYUKI HAEBARA  
WILLIAM E. HOFFMAN  
RAINER DIETZ  
PIERRE SCHELLING  
DETLEV GANTEN

Department of Pharmacology,  
University of Heidelberg,  
Im Neuenheimer Feld 366,  
6900 Heidelberg, FRG

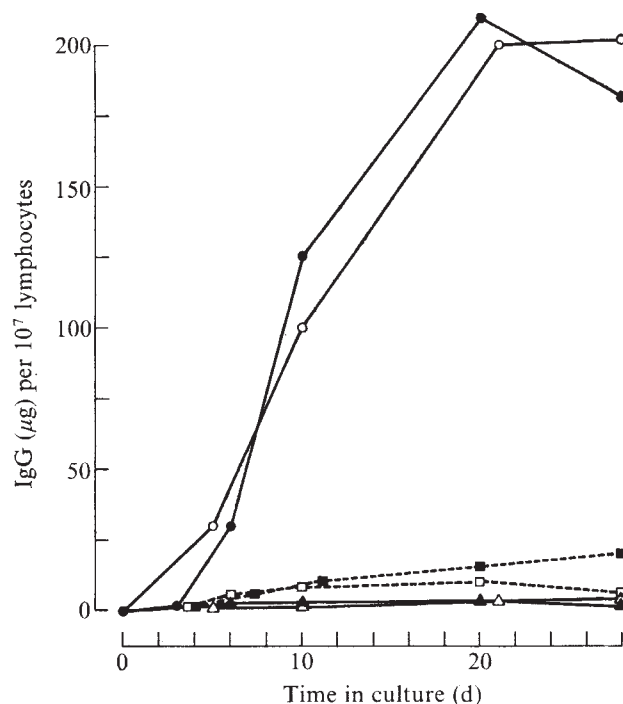
Received 27 July; accepted 27 August 1977.

\*To whom reprint requests should be addressed. Present address: Department of Physiology, University of Iowa, Iowa City, Iowa 52240 USA.

- Okamoto, K. (ed.) *Spontaneous Hypertension: Its Pathogenesis and Complications* (Igaku Shoin Ltd, Tokyo, 1972).
- Sen, S., Smeby, R. R. & Bumpus, F. M. *Circ. Res.* 31, 876-882 (1972).
- Koletzky, S., Shook, S. & Rivera-Velez, J. in *Spontaneous Hypertension: Its Pathogenesis and Complications* (ed. Okamoto, K.) 199-202 (Igaku Shoin Ltd, Tokyo, 1972).
- Ganten, D. *et al. Science* 173, 64-65 (1971).
- Fischer-Ferrario, C., Nahmod, V. E., Goldstein, D. J. & Finkelman, S. J. *exp. Med.* 133, 353-361 (1971).
- Ganten, D., Hutchinson, J. S., Schelling, P., Ganten, U. & Fischer, H. *Clin. exp. Pharmacol. Physiol.* 2, 103-126 (1976).
- Ganten, D. *et al. Am. J. Physiol.* 221, 1733-1737 (1971).
- Yang, H.-Y. T. & Neff, N. H. *J. Neurochem.* 21, 2443-2450 (1973).
- Severs, W. B. & Daniels-Severs, A. E. *Pharmacol. Rev.* 25, 415-449 (1973).
- Phillips, M. I. & Felix, D. *Brain Res.* 110, 531-540 (1976).
- Felix, D. *Naunyn-Schmiedeberg's Arch. Pharmacol.* 15, 292-295 (1976).
- Nicoll, R. A. & Barker, J. L. *Nature new Biol.* 172, 233-234 (1971).
- Bennet, J. P. & Snyder, S. H. *J. biol. Chem.* 251, 7423-7430 (1976).
- Ganten, D., Hutchinson, J. S. & Schelling, P. *Clin. Sci. molec. Med.* 48, 265s-268s (1975).
- Hoffman, W. E., Phillips, M. I. & Schmid, P. *Am. J. Physiol.* 232, H426-H433 (1977).
- Phillips, M. I., Phipps, J., Hoffman, W. E. & Leavitt, M. *Physiologist* 18, 350 (1975).
- Sweet, C. S., Colombo, J. C. & Gaul, S. K. *Fedn Proc.* 35, 1056 (1976).
- Schoelkens, B. A. *J.R.C.S. Med. Sci.* 4, 320-321 (1976).
- Elghozi, J. L. *et al. Clin. Sci. molec. Med.* 51, 385s-389s (1976).
- Fuxe, K., Ganten, D., Hökfelt, T. M. & Bolme, P. *Neurosci. Lett.* 2, 229-234 (1976).

## Thyroid-stimulating autoantibody production *in vitro*

GRAVES' disease is an autoimmune disorder in which hyperthyroidism is associated with the presence of thyroid-stimulating autoantibodies<sup>1</sup>. Recent work suggests that these antibodies, collectively termed TSAbs, are antibodies to the thyrotrophin (TSH) receptor which stimulate the thyroid by binding to the receptor and activating adenylate cyclase<sup>1-5</sup>. Synthesis of the autoantibody *in vivo* may result from defects in both the thyroid and the immune system. In the case of the thyroid, this may be an aberration in the synthesis or shedding of plasma membrane components related to the TSH receptor. In the case of the immune system, a current view is that autoantibody formation is the result of a suppressor T cell defect<sup>6</sup>. An *in vitro* model of TSAbs synthesis would permit an analysis of the role of the thyroid and the immune system in the events leading to thyroid-stimulating antibody production. Attempts have been made to study TSAbs production in short-term cultures of peripheral blood lymphocytes from Graves' patients.



**Fig. 1** Kinetics of IgG production by peripheral blood lymphocytes. Lymphocytes were separated on a Ficoll-Hypaque gradient<sup>19</sup> and washed three times; the yield from heparinised blood averaged 80%. The cells were incubated at  $1.0-1.5 \times 10^7$  per ml in 1 ml aliquots in Marbrook flasks with an outer reservoir of 25 ml. The culture medium was RPMI 1640 buffered with  $\text{NaHCO}_3$  and supplemented with L-glutamine (2 mM), penicillin ( $50 \text{ U ml}^{-1}$ ), streptomycin ( $50 \text{ µg ml}^{-1}$ ) and 10% heat inactivated foetal calf serum (Flow Laboratories, Lot 428136). Lymphocytes were incubated at  $37^\circ\text{C}$  in 5%  $\text{CO}_2$  in air for intervals of up to 28 d and collected as follows: the cell suspensions were removed from the Marbrook flasks, centrifuged and the culture supernatants separated from the cell pellets. Soluble material was extracted from the cell pellets by freezing, thawing and homogenising; the homogenates were centrifuged ( $100,000g$  for 30 min) and the supernatants (termed the soluble cell extract) collected. IgG was measured in the culture supernatants and soluble cell extracts by a solid phase radioimmunoassay<sup>20</sup>. Each kinetic curve is derived from the lymphocytes separated from 100 ml of blood from a single donor; each point is the mean of duplicate or triplicate cultures. Results are expressed as  $\text{µg IgG}$  produced per  $10^7$  cells initially added to each flask. ● and ○, IgG in the supernatants from cultures stimulated with PWM (Gibco,  $10 \text{ µl ml}^{-1}$ ); ▲ and △, IgG in cell extracts of cultures stimulated with PWM; ■ and □, IgG in the supernatants of cultures without mitogen.

These studies, however, did not use optimal conditions for immunoglobulin production and the levels of TSAbs reported were close to the limits of detectability of the imprecise assay methods used<sup>7-11</sup>. In the study described here we have used the optimal conditions for culturing lymphocytes provided by Marbrook flasks<sup>12</sup> in conjunction with two new sensitive and precise methods for detecting TSAbs—the radio-receptor<sup>13</sup> and cytochemical assays<sup>14,15</sup>. With this system we have readily been able to detect TSAbs production in cultures of Graves' lymphocytes.

Thyroid-stimulating antibodies are always associated with the IgG fraction of serum. Therefore in preliminary experiments a study of the kinetics of IgG production *in vitro* was made using peripheral blood lymphocytes from normal donors. Lymphocytes were cultured for up to 28 d in medium alone or medium containing pokeweed mitogen (PWM) and the IgG in the culture supernatants measured by solid phase radioimmunoassay. Also, the amount of immunoglobulin remaining in the cells was determined. Some representative experiments are shown in Fig. 1. After 10 d of culture in the presence of PWM the medium contained about  $100 \text{ µg}$  of IgG per  $10^7$  cells and this was increased to  $200 \text{ µg}$  after 21 d. The cell extracts contained far less IgG. In the absence of PWM, only small

**Table 1** TSAb activity in supernatants from lymphocyte cultures measured by cytochemical assay

| Culture conditions | Patient | IgG dose ( $\mu$ g) | % Stimulation | Donor | IgG dose ( $\mu$ g) | % Stimulation |
|--------------------|---------|---------------------|---------------|-------|---------------------|---------------|
| Membranes + PWM    | Bu      | 0.64                | 35.7          | Bo    | 0.22                | 0.4           |
| PWM                | Ca      | 0.88                | 28.6          | Pe    | 0.60                | 2.9           |
| PWM                | GI      | 0.82                | 38.0          | Sm    | 0.56                | 2.2           |

Lymphocytes were incubated with PWM, alone or after an initial 24-h period with thyroid membranes. Immunoglobulins in the culture supernatants and medium alone were precipitated using ammonium sulphate, dialysed exhaustively against isotonic saline and assayed in the section assay for thyroid stimulators<sup>21</sup>. The results are expressed as % stimulation of the absorbance ( $A$ ) over control values (medium only) calculated from the formula:

$$[(A_{\text{test sample}} - A_{\text{control}})/A_{\text{control}}] \times 100$$

Each absorbance value is the mean of 10 measurements per slide on duplicate slides. The amount of IgG used for each assay is given. All three Graves' patients studied had high levels of serum TSAb activity.

amounts of IgG were detected in the medium and cell extracts. The ability of PWM to stimulate immunoglobulin production in human lymphocyte cultures is well known<sup>16,17</sup>. On the basis of these experiments it was decided to investigate thyroid-stimulating antibody production by culturing Graves' lymphocytes with PWM and assaying the culture supernatants for TSAb after 14 d. This time interval was chosen so as to provide near-maximal amounts of IgG while reducing the risk of infection which is a problem with long-term cultures.

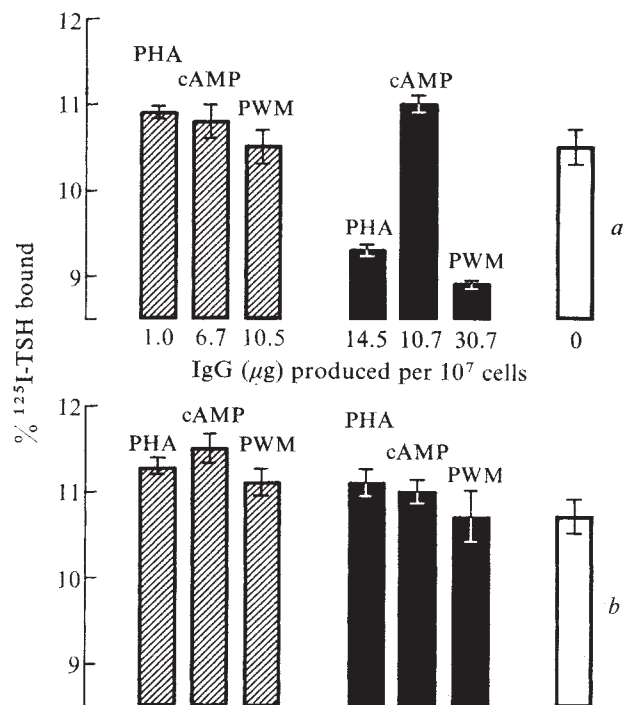
Table 1 shows the measurement of TSAb activity by cytochemical assay in the immunoglobulin fraction of the culture supernatants from Graves' and normal peripheral blood lymphocytes. The assay depends on detecting TSAb- (or TSH)-induced increases in the permeability of thyroid cell lysosomes. These increases are monitored by using a chromogenic substrate (leucyl- $\beta$ -naphthylamide) for the lysosomal enzyme naphthylamidase and the intensity of the resulting intracellular colour is quantitated by scanning and integrating microdensitometry. The supernatants from all three Graves' lymphocyte cultures produced an increase in absorbance of 30–40% over control (medium only) whereas the responses elicited by supernatants from cultures of normal lymphocytes were less than 3%. This study, in which the assay of the supernatants from cultures of normal and Graves' lymphocytes was carried out 'blind', provided good evidence that TSAb was produced by the cultured Graves' lymphocytes.

Confirmation of the presence of TSAb in Graves' lymphocyte cultures was provided by studies with the radioreceptor assay for thyroid-stimulating antibodies which depends on the ability of TSAb to inhibit the binding of labelled TSH to thyroid membranes<sup>1,3-5,13</sup>. In these experiments, lymphocytes were cultured with antigen (in the form of thyroid membranes) for 24 h. One-third of the cells were also treated with dibutyl cyclic AMP during this period since there is evidence from experiments with mice that cyclic AMP administered with antigen *in vivo* and *in vitro* results in higher numbers of antibody-forming cells<sup>18</sup>. After 24 h, the cells were washed and resuspended in medium alone (for cyclic AMP-treated cells) or medium plus PWM or phytohaemagglutinin (PHA). The supernatants from 14-d cultures were assayed for IgG and the immunoglobulins of the supernatants monitored for TSAb activity in the receptor assay. A significant inhibition of labelled TSH binding to thyroid membranes was observed with supernatants from Graves' lymphocytes cultured with PHA (Student's  $t = 24.5$ ,  $P < 0.001$ ) and PWM ( $t = 7.1$ ,  $P < 0.01$ ) and this indicated the presence of TSAb (Fig. 2a). The specificity of the effect of the cultures in the receptor assay was demonstrated by the experiment shown in Fig. 2b in which supernatants from the same cultures were pre-incubated with thyroid membranes before assay; the TSAb activity originally present in the PHA and PWM cultures was absorbed out.

These observations indicated that peripheral blood lymphocytes from patients with Graves' disease are capable of making thyroid-stimulating autoantibodies when cultured under the conditions provided by Marbrook flasks<sup>22</sup>. It should now be possible to use the combination of the Marbrook culture

system with the precise and sensitive cytochemical and radio-receptor assays to investigate the events leading to TSAb production. Similarly, autoantibody production in other human autoimmune diseases could be studied using this approach.

We thank Drs Chayen and Bitensky, from the Kennedy Institute of Rheumatology, for considerable help with the cytochemical bioassay. We also thank Drs J. G. Pierce and J. Fawcett and the Armour Pharmaceutical Company for gifts of TSH, Professor J. Owen, Professor A. L. Latner, Professor



**Fig. 2** TSAb activity in supernatants of lymphocyte cultures measured by radioreceptor assay<sup>13</sup>. Lymphocytes were incubated with thyroid membranes (35  $\mu$ g protein  $\text{ml}^{-1}$ ) for 24 h, washed and then cultured with PHA (Wellcome 10  $\mu\text{l ml}^{-1}$ ), PWM (10  $\mu\text{l ml}^{-1}$ ) or medium only for 14 d; cells receiving medium only had been stimulated with dibutyl cyclic AMP (cAMP, 0.4 mM) during the 24 h incubation with membranes. IgG was measured in the culture medium and the amount produced per  $10^7$  cells is shown. Immunoglobulins were precipitated from the culture supernatants with ammonium sulphate, dialysed exhaustively against 50 mM NaCl, 10 mM Tris-HCl (pH 7.4) and assayed in the radioreceptor assay. The figure shows the effect of the immunoglobulins on <sup>125</sup>I-TSH binding to thyroid membranes. *a*, Immunoglobulins assayed directly. *b*, Immunoglobulins pre-incubated with human thyroid membranes. All values are the means of triplicate  $\pm$  s.e.m. Data are given for a Graves' patient (solid bars), a normal donor (cross-hatched bars) and culture medium alone (open bars). The same number of cells were cultured from the Graves' patient and the normal donor. The Graves' patient studied in this experiment had particularly high levels of serum TSAb; 1  $\mu$ g of purified serum IgG was readily detectable in the receptor assay.



I. M. Roitt, Drs J. North, T. Platts-Mills, P. Lydyard and N. Wilcox for useful discussion and Mrs Alison Scaife for technical assistance. The work was supported by grants from the MRC, the North of England Campaign for Cancer Research, the Wellcome Trust, the League of Friends of the Royal Victoria Infirmary, the Ernest and Minnie Dawson Cancer Trust and the Scientific and Research Committee of the Newcastle University Hospitals.

S. M. McLACHLAN  
B. REES SMITH  
V. B. PETERSEN  
T. F. DAVIES  
R. HALL

Departments of Medicine and Clinical Biochemistry,  
University of Newcastle upon Tyne,  
Newcastle upon Tyne, UK

Received 2 September; accepted 30 September 1977.

- Hall, R., Rees Smith, B. & Mukhtar, E. D. *Clin. Endocr.* **4**, 213–230 (1975).
- Mukhtar, E. D., Rees Smith, B., Pyle, G. A., Hall, R. & Vice, P. *Lancet* **i**, 713–720 (1975).
- Fayet, G. *et al.* *FEBS Lett.* **32**, 299–302 (1973).
- Manley, S. W., Bourke, J. R. & Hawker, R. W. *J. Endocr.* **61**, 437–445 (1974).
- Mehdi, S. Q. & Nussey, S. S. *Biochem. J.* **145**, 105–111 (1975).
- Allison, A. C. & Denman, A. M. *Br. med. Bull.* **32**, 124–129 (1976).
- McKenzie, J. M. & Gordon, J. in *Current Topics in Thyroid Research* (eds Cassano, C. & Andreoli, M.) 445 (Academic, New York, 1965).
- Miyai, K., Fukuchi, M., Kumahara, U. & Abe, H. *J. clin. Endocr. Metab.* **27**, 855–860 (1967).
- Wall, J. R., Good, B. F., Forbes, I. J. & Hetzel, B. S. *Clin. exp. Immun.* **4**, 555–561 (1973).
- Knox, A. J. S., von Westarp, C., Row, V. V. & Volpé, R. *J. clin. Endocr. Metab.* **43**, 330–337 (1976).
- Knox, A. J. S., von Westarp, C., Row, V. V. & Volpé, R. *Metabolism* **25**, 1217–1223 (1976).
- Marbrook, J. *Lancet* **ii**, 1279–1281 (1967).
- Rees Smith, B. & Hall, R. *Lancet* **ii**, 427–431 (1974).
- Bitensky, L., Alaghband-Zadeh, J. & Chayen, J. *Clin. Endocr.* **3**, 363–374 (1974).
- Petersen, V., Smith, B. R. & Hall, R. *J. clin. Endocr. Metab.* **41**, 199–202 (1975).
- Wu, L. Y. F., Lawton, A. R. & Cooper, M. D. *J. clin. Invest.* **52**, 3180–3189 (1973).
- Janossy, G., Gomez de la Concha, E., Waxdal, M. J. & Platts-Mills, T. *Clin. exp. Immun.* **26**, 108–117 (1976).
- Weinstein, Y. & Melmon, K. L. *Immun. Comm.* **5**, 401–416 (1976).
- Boyüm, A. *Scand. J. Immun.* **5** suppl. 5, 9–15 (1976).
- Wide, L., Nilius, S. J., Gemzell, C. & Roos, P. *Acta Endocr. suppl.* **174**, 1–58 (1974).
- Gilbert, D. M., Besser, G. M., Bitensky, L. & Chayen, J. *J. Endocr.* (in the press).
- North, J. R. & Maizeis, R. M. *Immunology* **32**, 771–776 (1977).

## Relaxin has conformational homology with insulin

RELAXIN, a polypeptide hormone synthesised and stored in the corpus luteum, is responsible for the dilation of the symphysis pubis in most mammals before parturition. Porcine relaxin (molecular weight ~5,600) consists of one A chain and one B chain linked by disulphide bonds. The amino acid sequence of the two chains is consistent with interchain and intrachain disulphide crosslinks of the same disposition as those of insulin<sup>1,2</sup>. Although only five further residues are identical to those in equivalent positions of porcine insulin, the model presented here shows that relaxin may have a tertiary structure closely resembling that of

insulin. The residues of the hydrophobic core of relaxin differ from those of insulin but are close packed and occupy the same volume.

Figure 1 shows the sequence of porcine relaxin at determined by Schwabe *et al.*<sup>1,2</sup> aligned with the sequence of porcine insulin according to the positions of the cystines (for reviews see refs 3, 4). In this paper the insulin numbering scheme has been applied to the relaxin sequence to facilitate comparison of the equivalent residues in the two molecules. Some relaxin molecules apparently have additional residues, possibly serine and arginine at the C-terminus of the B chain<sup>2</sup>. This may, by analogy with insulin, be a consequence of there being a precursor, prorelaxin; these derivatives may be intermediates in the conversion process. A further sequence analysis of a porcine relaxin by Niall and co-workers<sup>5</sup> shows that there may be some heterogeneity in relaxin. Since these differences seem to be limited to residues at the B-chain termini located on the outside of the model presented here, they could be accommodated without affecting the relaxin tertiary structure that we describe.

We have explored a possible three-dimensional structural homology of insulin and relaxin in the following way. A model was built first of the main chain A1–A20 and B6–B21, in which the interchain and intrachain disulphide bridges occupied equivalent three-dimensional positions to those reported for insulin, and other amino acid residues shown to be equivalent in the sequence alignment of Fig. 1 were given the same main chain conformation. This was made possible by the presence of glycines at B8 and B20 in relaxin as in insulin, in which positions an amino acid with a side chain would not be allowed as a consequence of the positive  $\psi$  angles. The side chains of B6, B11, B12, B14, B15, B18, A2 and A16 which contribute to the hydrophobic core of insulin were then built into the framework provided by the main chain and cystine disulphides. These are all hydrophobic in relaxin and are easily accommodated to give a close packed core. B11 and B12 are conserved as in all insulins. Of particular interest is the complementary nature of many of the sequence differences. For example, in insulin B6 Leu and B14 Ala are in close juxtaposition; in relaxin their respective positions are reversed: B6 Ala and B14 Leu. In a similar way A2 Ile and A16 Leu of insulin are A2 Leu and A16 Ile in relaxin. These pairs of side chains point towards the centre of the core from opposite sides. The close-packed occupancy of the core is thus maintained. The B15 Trp of relaxin is beautifully accommodated almost completely buried in the core with no relative changes of the positions of the A and B chains. The proposed main chain conformation of relaxin is shown in Fig. 2 and a stereoview of the three-dimensional structure is shown in Fig. 3.

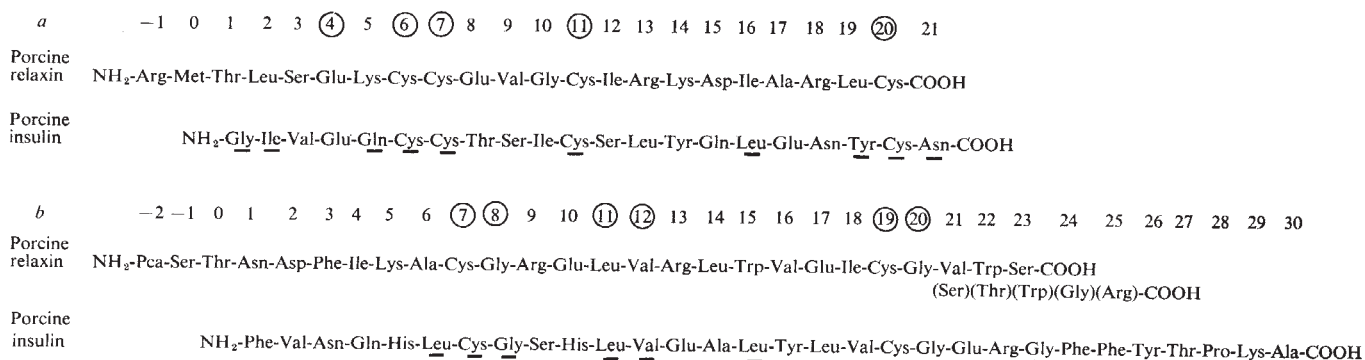
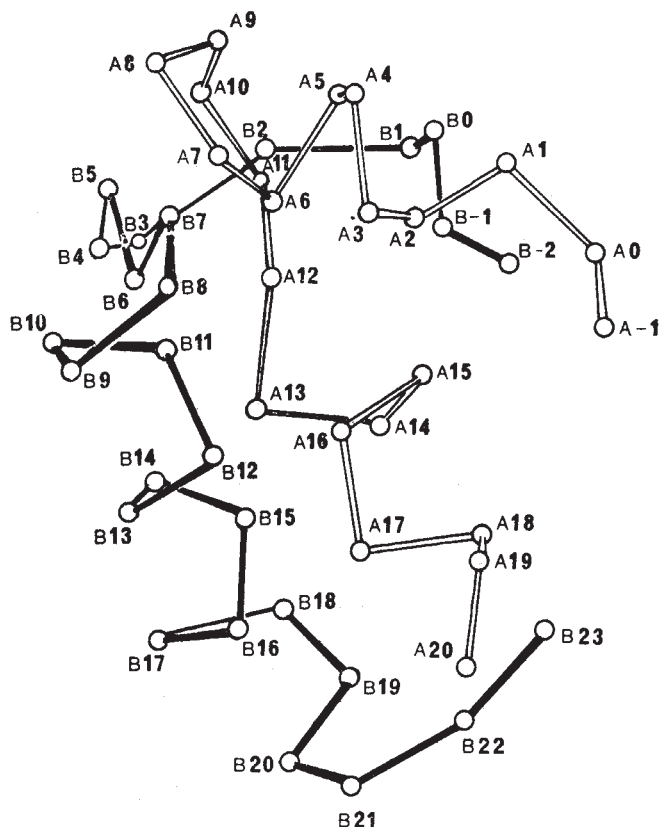


Fig. 1 Sequences of porcine relaxin<sup>1,2</sup> and insulin arranged to give alignment of the cystines. There may be further Ser and Arg residues at B24 and B25 in some relaxin molecules<sup>2</sup> and an independent sequence determination<sup>5</sup> indicates possible heterogeneity at B-2, B21, B22 and B23 (the sequence differences are given in parentheses). The numbers of residues identical in porcine insulin and relaxin are circled and residues invariant in all known insulins are underlined.





**Fig. 2** Proposed conformation of the relaxin main chain indicated by  $\alpha$  carbon positions ( $\circ$ ) and joined by virtual bonds. The residue numbers are those shown in Table 1 to facilitate comparison with insulin. The molecule is viewed from a position equivalent to right angles to the three-fold axis and into the face involved in dimerisation of the insulin hexamer (see Fig. 10 of ref. 3). In insulin the B chain continues from B23 to a position close to A2 partly covering the A chain residues as viewed here.

Having established the feasibility of an insulin-like fold for relaxin we then proceeded to add surface residues including the chains B6–B22 and A1–A20. This naturally led to the suggestion of many ion pairs in particular A5 Lys and A15 Asp of relaxin, which are A5 Gln and B15 Gln in insulin. Further possible relaxin ion pairs are A8 Glu and B5 Lys, B13 Arg and B17 Glu.

Residue A19 is Tyr in insulin and Leu in relaxin. This change is easily accommodated in the model and may be related to the extra residues at the A-chain N terminus of

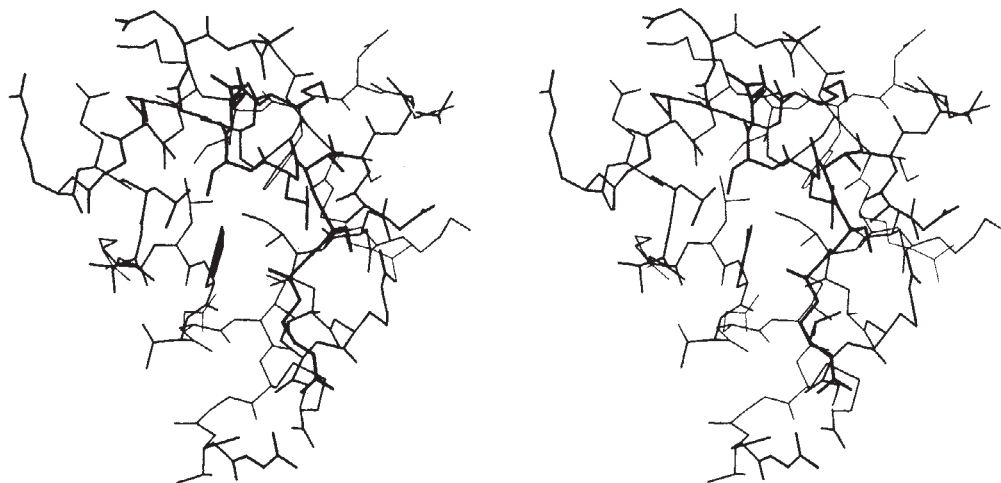
relaxin which are most probably covering A19 so that A0 Met is in contact with Leu, leading to a further ion pair A–1 Arg to the B-chain C terminus. In a similar way the extra residues at the B-chain N terminus are easily accommodated. When residue B3 Phe was allowed to lie against the residues of the hydrophobic core, the chain could be further extended to fold over A12, a position which is Ile in relaxin although always hydrophilic in insulins. But, the conformations of the B-chain N terminal residues may be flexible in relaxin as they are in insulin, and it is possible that they contribute little to the stability of the insulin-like core. The absence of a residue at A21 might be expected to destabilise the insulin-like tertiary structure, but would not prevent its existence.

Although of less consequence for the tertiary structure of relaxin, the positioning of the residues at the B-chain C terminus must await clarification of this part of the primary structure. If the polypeptide chain has either of the reported sequences<sup>2,5</sup>, it can be folded back on to the hydrophobic core. The tryptophan probably contributes towards stabilising the tertiary structure. This residue may be placed so that it partially compensates for the lack of residues in relaxin equivalent to B24 Phe and B26 Tyr of insulin. Thus the tertiary structure would be expected to be more stable than that of desoctapeptide insulin (B23–B30 deleted) which has a tertiary structure rather different from that of native insulin. Residue A3 in our model of relaxin—a serine residue—occurs on the surface of the molecule whereas in insulin this residue is usually a valine and is close to the B-chain C terminus.

The uncertainty in the C-terminal B-chain sequence may be due to the existence of two variants of relaxin suggesting that this portion of the molecule makes only a limited, contribution, if any, to the biological activity. Thus, the tryptophan residue known to be inessential<sup>6</sup> is most likely the one located in the C-terminal tail of the relaxin B chain. Our model predicts that the remaining tryptophan B15 which is buried cannot be oxidised in native relaxin without loss of tertiary structure and probably therefore biological activity. The relative mobility of the C-terminal peptide of the B-chain may be deduced from the fact that carboxypeptidase A digestion of native relaxin readily releases Ser, Trp, Val, and somewhat more slowly, Gly.

The absence of histidine at B10 implies that relaxin will not bind zinc ions to give a hexamer in the same way as insulin, although the B10 Glu of relaxin may bind other metal ions such as calcium. But, the association of the protomers would also be hindered by the change of the C-terminus of the relaxin B-chain and the presence of charged groups at B17 and A13 in relaxin which are usually hydrophobic in insulin. The equivalent groups in

**Fig. 3** A stereo view of the relaxin model from approximately the same direction as Fig. 2. All side-chain residues are included with the exception of those of B21, B22 and B23. The tryptophan (B15) is seen centrally end on and partially buried in the hydrophobic core. The drawing is taken from coordinates measured from the model and constrained towards the known geometry of amino acids using the method described in ref. 9.



insulin are involved in protomer-protomer interactions in the zinc insulin hexamer.

The differences between residues at A1 and A19 and the deletions at the two chain termini make it unlikely that relaxin will bind insulin receptors, since modification or deletion of these residues in insulin leads to considerable loss of affinity for the receptor and a decrease of biological activity<sup>3,7</sup>.

Relaxin in its highly purified form requires an adjuvant such as beeswax or benzopurpurin 4- $\beta$  for enhancement of biological activity. This may well be due to interaction of the diazo dye, for example, with the hydrophobic residues, particularly tryptophan B22 and subsequent stabilisation of the structure and protection from proteolytic attack. The induction of a strong Cotton effect at 520 nm indicates that the dye molecules bind relaxin strongly and specifically (data not shown).

The existence of a molecule with the sequence of relaxin and an insulin-like tertiary structure may be explained by gene duplication of primitive insulin-like molecules followed by accepted mutations leading to radical changes in sequence<sup>8</sup>. If this is so, the nature of the complementary changes in the hydrophobic core implies evolution through a less easily folded state possibly implying a period without selective pressure on the duplicated gene.

We thank Mr D. Richardson and Dr P. Pauling of University College, London, and Mr T. Sewell and Mr J. Jenkins for assistance with computer graphics, Dr J. L. Finney and Mr J. D. Nicholas for use of their coordinate measuring device, and Miss S. Kennett for assistance with photography of the model. T.L.B. and S.B. are grateful to the SRC for support.

*Note added in proof:* An independent model building study using computer graphics has also concluded that relaxin has an insulin-like conformation (N. Isaacs *et al.*, manuscript in preparation).

S. BEDARKAR  
W. G. TURNELL  
T. L. BLUNDELL

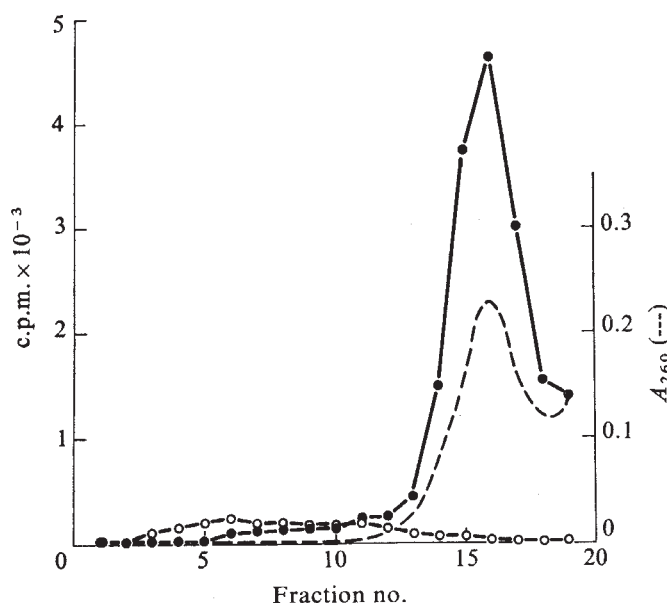
Laboratory of Molecular Biology,  
Department of Crystallography,  
Birkbeck College, University of London,  
Malet Street, London WC1, UK

C. SCHWABE

Department of Biochemistry,  
Medical University of South Carolina,  
Charleston, South Carolina 29401

Received 6 July; accepted 4 October 1977.

- Schwabe, C., McDonald, J. K. & Steinetz, B. G. *Biochem. biophys. Res. Commun.* **70**, 397-405 (1976).
- Schwabe, C., McDonald, J. K. & Steinetz, B. G. *Biochem. biophys. Res. Commun.* **75**, 503-510 (1977).
- Blundell, T. L. *et al. Adv. Protein Chem.* **26**, 280-394 (1972).
- Blundell, T. L. & Wood, S. P. *Nature* **257**, 197-203 (1975).
- Kwok, S., Bryant-Greenwood, G., James, R. & Niall, H. *Nature* **267**, 544-546 (1977).
- Schwabe, C. & Braddon, S. A. *Biochem. biophys. Res. Commun.* **68**, 1126-1132 (1976).
- Pullen, R. A. *et al. Nature* **259**, 369-373 (1976).
- Schwabe, C. & McDonald, J. K. *Science* (in the press).
- Dodson, E. J., Isaacs, N. W. & Rollett, J. S. *Acta crystallogr.* **A32**, 311-315 (1975).



**Fig. 1** Repair replication in confluent human fibroblasts. Cultures were incubated in the presence of BUdR  $10^{-5}$  M and FUdR  $2 \times 10^{-6}$  M for 1 h, irradiated with  $13 \text{ J m}^{-2}$  ultraviolet light and labelled for 2 h in BUdR  $10^{-5}$  M,  $^3\text{H}$ -Tdr  $10 \mu\text{Ci ml}^{-1}$  (64 Ci mmol $^{-1}$ ), FUdR  $2 \times 10^{-6}$  M, hydroxyurea  $2 \times 10^{-3}$  M; the DNA was isolated and sedimented in alkaline caesium chloride caesium sulphate gradients<sup>10,11</sup>. The exclusive location of  $^3\text{H}$  label at normal density (coincident with the absorbance peak) indicates that the  $^3\text{H}$  has been incorporated by repair replication. Broken line,  $A_{260}$ ; ○, controls; ●, irradiation with  $13 \text{ J m}^{-2}$  ultraviolet. Radioactivity profiles were normalised to the same amount of DNA per gradient.

as discrete particles<sup>2</sup>. The structural organisation of mammalian DNA into nucleosomes, and higher degrees of order in the packing of strings of nucleosomes into chromosomal fibres has been shown to play a major part in the control of DNA replication<sup>4,5</sup>, and transcription<sup>6</sup>. Previous studies have indicated that the accessibility of damage in DNA to repair enzymes is restricted by chromatin proteins<sup>7,8</sup> and it is therefore likely that the repair of damaged sites in DNA is also controlled by some features of nucleosome structure and packing<sup>9</sup>. In the study described here I have measured the rates of degradation of repaired regions in mammalian DNA by micrococcal nuclease; the results suggest that the first sites of ultraviolet light-induced damage to be repaired are those in the DNA between nucleosomes and that there is little rearrangement of nucleosomes along the DNA during repair.

Human fibroblast cultures (strain E11) were established from foreskin and grown in Eagle's minimal essential medium with 10% foetal calf serum. Cultures for experiments were grown in  $0.01 \mu\text{Ci ml}^{-1}$ ,  $56.4 \text{ mCi mmol}^{-1}$   $^{14}\text{C}$ -thymidine ( $^{14}\text{C}$ -Tdr) in 90-mm Petri dishes until confluent and mitotic figures no longer detectable. Cultures were then rinsed and grown for 24 h in unlabelled medium before being once irradiated with  $13 \text{ J m}^{-2}$  ultraviolet light (254 nm, at a dose rate of  $1.3 \text{ J m}^{-2} \text{ s}^{-1}$ ). They were then labelled with  $^3\text{H}$ -Tdr ( $10 \mu\text{Ci ml}^{-1}$ , 64 Ci mmol $^{-1}$ ), hydroxyurea ( $5 \times 10^{-3}$  M), and fluorodeoxyuridine (FUdR,  $2 \times 10^{-6}$  M) for 10 min to 4.5 h. For most experiments cells were collected immediately but some were allowed to grow in  $5 \times 10^{-4}$  M Tdr and  $5 \times 10^{-5}$  M deoxycytidine for several hours before collection.

Confluent cultures of primary fibroblasts perform negligible amounts of semiconservative DNA replication, and all of the radioactivity incorporated into DNA in the presence of hydroxyurea after ultraviolet irradiation consists of repair replication (Fig. 1). Therefore, in cultures allowed to reach confluency in  $^{14}\text{C}$ -Tdr, irradiated with ultraviolet light and labelled with  $^3\text{H}$ -Tdr, the  $^{14}\text{C}$  activity is uniformly distributed throughout the DNA but the  $^3\text{H}$  activity is confined to the short patches which are involved in repair of ultraviolet damage in DNA<sup>10,11</sup>.

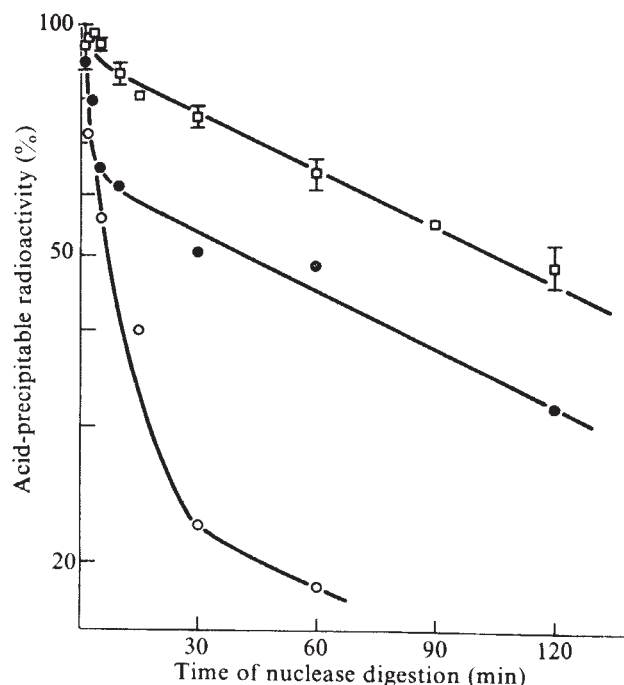
## Nucleosome structure controls rates of excision repair in DNA of human cells

EUKARYOTIC chromatin is now considered to be composed of structural units, nucleosomes (nu bodies), which consist of four pairs of histones around which DNA is coiled<sup>1-3</sup>. Approximately 200 base pairs of DNA are associated with each nucleosome, 140 base pairs are wrapped around each and a variable length of approximately 40 base pairs lies between<sup>1-3</sup>. The DNA wrapped around each nucleosome is less easily digested by micrococcal nuclease than the DNA between them, and brief periods of incubation with the enzyme allow isolation of nucleosomes

Repaired regions in cells that had incorporated  $^3\text{H}$ -Tdr immediately after ultraviolet irradiation were more rapidly digested by nuclease action than was  $^{14}\text{C}$ -Tdr-labelled DNA (Fig. 2, Table 1). When cells that had incorporated  $^3\text{H}$ -Tdr into repaired DNA immediately after irradiation had grown for 5 h, the  $^3\text{H}$  label was digested to a similar extent to that in cells collected earlier (Fig. 2, Table 1). When cells were labelled for longer periods the proportion of  $^3\text{H}$  activity that was rapidly digested decreased (Fig. 2, Table 1).

These results show that the rate at which nuclease degrades  $^3\text{H}$ -labelled DNA in repaired regions is different from that at which  $^{14}\text{C}$  uniformly labelled DNA is degraded. To a first approximation, the degradation kinetics can be analysed in terms of two exponential components (Fig. 2): an initial rapid component which is present in decreasing proportion as the labelling time increases, and a slow component which has approximately the same slope at all labelling times (Fig. 2, Table 1). The simplest interpretation of these results is that the rapid component represents degradation of the DNA between nucleosomes, and the slow component represents degradation of DNA closely associated with nucleosomes. This interpretation is strengthened by observation that the DNA after 5 min digestion with nuclease in essentially the same conditions in this laboratory is in one or more multiples of approximately 200 nucleotide pair lengths (B. Young, unpublished). The results indicate therefore, that at short times after irradiation the repaired regions are predominantly between nucleosomes, but after longer times all regions of DNA contain repaired regions.

There are two possible reasons for the change in the distribution of  $^3\text{H}$  label in repaired regions with increasing labelling times. One is that only internucleosome DNA is repaired and that there is a redistribution of nucleosomes along the DNA which tends to randomise the label. The other is that internucleosome DNA is more accessible to repair enzymes and is repaired more rapidly than DNA on the nucleosomes. Because the inter-nucleosome DNA and hence the ultraviolet-induced photoproducts in this DNA are a minor fraction of the whole DNA and a minor fraction of the total repair replication is completed in the first hour after irradiation (Table 1), faster repair in these regions will result in earlier completion of repair; at longer times repair on nucleosomes would predominate. The first possibility is eliminated, however, by the pulse-chase experiments. Growing cells for 5 h after a 30-min pulse of  $^3\text{H}$ -Tdr immediately following irradiation produced a  $^3\text{H}$  distribution that was similar to that obtained from similarly labelled cells collected earlier (Table 1). Cells that had been labelled for only 10 min, however, had a  $^3\text{H}$  distribution in which the rapidly digested component decreased from 58% to 39%. These observations at 10 and 30 min of labelling indicate that the distribution of DNA between and on nucleosomes changes little during 5 h of growth. The slight change observed in the 10-min labelled sample could actually be due to variability



**Fig. 2** Percentage of  $^{14}\text{C}$ - or  $^3\text{H}$ -labelled DNA in nuclei remaining undigested by micrococcal nuclease after various times of incubation at  $37^\circ\text{C}$ . After cultures were labelled and irradiated as described in the text, cells were collected by rinsing once in physiological buffer, trypsin treatment for 2 min at  $37^\circ\text{C}$  and the suspension was centrifuged at  $1,800g$  for 1 min. Cells were resuspended in ice-cold hypotonic buffer (RSB,  $0.0015\text{ M MgCl}_2$ ,  $0.015\text{ M NaCl}$ ,  $0.01\text{ M Tris-HCl}$ ,  $\text{pH } 7.4$ ) and centrifuged again for three successive washes. They were resuspended in RSB containing  $1\%$  Triton X-100 and gently homogenised by 10 strokes in a tight fitting Dounce homogeniser to prepare nuclei<sup>5</sup>. Nuclei were centrifuged for 1 min at  $1,800g$  and then resuspended in RSB and centrifuged three more times. The nuclear pellet was resuspended in RSB,  $\text{CaCl}_2$  added to a final concentration of  $10^{-3}\text{ M}$  and the suspension dispensed in  $0.4\text{ ml}$  quantities into tubes containing  $2\text{ }\mu\text{l}$  of micrococcal nuclease (Worthington, 40 units of enzyme to approximately  $10^4$  nuclei), and incubated at  $37^\circ\text{C}$ . Individual tubes were removed at various times, chilled on ice,  $100\text{ }\mu\text{g}$  carrier DNA added and the DNA precipitated with perchloric acid (PCA) at a final concentration of  $5\%$ . The supernatant was counted in  $7\text{ ml}$  water miscible liquid scintillation fluid (Aquasol). The precipitate was dissolved in  $5\%$  PCA at  $90^\circ\text{C}$  and counted in Aquasol. The fractions of  $^{14}\text{C}$  and  $^3\text{H}$  radioactivity rendered acid soluble by micrococcal nuclease was calculated from the radioactivity in supernatant and precipitate fractions from each tube.  $\square$ , Mean and standard error of  $^{14}\text{C}$ -Tdr uniformly-labelled cells.  $\circ$ , Labelled for 10 min with  $^3\text{H}$ -Tdr immediately after  $13\text{ J m}^{-2}$  ultraviolet light.  $\bullet$ , Labelled for 30 min immediately after  $13\text{ J m}^{-2}$  ultraviolet light.

**Table 1** Percentage of total radioactivity in the rapidly digested components of DNA in nuclei digested with micrococcal nuclease

| Labelling and irradiation conditions                    | % Rapidly digested* |      | Repair replication† |
|---------------------------------------------------------|---------------------|------|---------------------|
|                                                         | 10                  | (10) |                     |
| Uniformly labelled $^{14}\text{C}$ -Tdr                 |                     |      | —                   |
| Repaired, 10 min, immediately post UV                   | 58                  | (2)  | 4                   |
| Repaired, 30 min, immediately post UV                   | 33                  | (2)  | 14                  |
| Repaired, 30 min, immediately post UV‡                  | 29                  | (1)  | 14                  |
| Repaired, 90 min, immediately post UV                   | 19                  | (2)  | 53                  |
| Repaired, 4.5 h, immediately post UV                    | 21                  | (2)  | 120                 |
| Repaired, 30 min, immediately post UV and grown for 5 h | 35                  | (1)  | —                   |
| Repaired, 10 min, immediately post UV and grown for 5 h | 39                  | (1)  | —                   |

\*Values were calculated by extrapolation of the slow component of parallel curves similar to those shown in Fig. 2. Each value was obtained from the average of at least two separate determinations and is therefore not necessarily the same as the values that can be read off Fig. 2 since the latter illustrates single experimental determinations at each labelling time (the number in parentheses indicates number of separate degradation curves used to estimate this percentage).

†Amounts of repair replication expressed in arbitrary units calculated from experiments similar to those of Fig. 1, done for various labelling times.

‡Labelling done with  $^3\text{H}$ -Tdr but without hydroxyurea and FUdr.  
UV, ultraviolet irradiation.



in the precise labelling duration, because the nuclease digestion kinetics change rapidly during the shorter labelling periods. The distribution of repaired regions is therefore probably due to a faster rate and earlier completion of repair in inter-nucleosome DNA than in nucleosomal DNA.

In a previous study of the repair of alkylation damage in mouse cells Bodell<sup>12</sup> also showed that repair was concentrated on inter-nucleosomal DNA. In that study, however, it could be argued that there was either a non-uniform distribution of damage or a non-uniform distribution of repair, or both, because alkylating agents do not react with all regions of DNA to equal extents<sup>13</sup>. In the present study on the repair of ultraviolet damage it is most likely that the non-uniformity I have observed is due to the mechanism of repair, because pyrimidine dimers are caused by direct absorption of quanta of ultraviolet light in pyrimidine rings and the yield of photoproducts is insensitive to DNA conformation<sup>14</sup>. Some small effect of DNA structure on dimer formation which would result in a non-random distribution of dimers within or between nucleosomes cannot be completely excluded, however, but is unlikely.

These results (Fig. 2) can be interpreted to indicate that damaged DNA between nucleosomes is more rapidly repaired than DNA on the nucleosomes, this would explain many features of DNA repair previously observed. The size of the patch that replaces excised pyrimidine dimers, for example, has been usually estimated after long labelling times<sup>15, 16</sup>, after which most of the repair would be in regions of DNA on nucleosomes. Patch sizes estimated by a variety of methods seem to be about half the single-strand length of DNA associated with one nucleosome<sup>9, 10</sup>. This is what would be predicted if a randomly located dimer in a nucleosome is the starting point for an excision repair patch and the next nucleosome is the termination point. This model also predicts that the first excision repair patches made between nucleosomes after ultraviolet irradiation should have a much smaller patch size than those made later.

Several previous observations are consistent with a model in which excision repair is controlled by the structure of chromatin. Paterson *et al.*<sup>17</sup> found that excision of dimers occurred at two rates, an initial fast rate and a later slow one, and many studies on repair replication showed similar rate changes<sup>18–20</sup>. The possibility that dimers are not all equally accessible to repair enzymes was suggested by Wilkins and Hart<sup>7</sup> who showed that those dimers that were more rapidly excised from DNA were also more accessible to an exogenously supplied *Micrococcus luteus* 'ultraviolet specific' endonuclease. Mortelmans *et al.*<sup>8</sup> also showed that several complementation groups of the human DNA repair deficient disease xeroderma pigmentosum lack cofactors that control the repair of DNA in chromatin, rather than lack repair enzymes themselves as originally thought<sup>15, 21</sup>.

I conclude from these results (Fig. 2, Table 1), that the structure of mammalian chromatin exerts restraints on the rate of excision repair of ultraviolet-induced damage. In constructing models for excision repair it should, therefore, be considered that perhaps the substrate itself is the more complex entity rather than the association of repair enzymes, which has been frequently suggested<sup>22–24</sup>. The need for controlled access to damage in DNA may be one reason for the greater complexity of repair mechanisms in eukaryotes than in prokaryotes, which is indicated by the multiple complementation groups which affect excision-repair in xeroderma pigmentosum<sup>25, 26</sup>.

I thank Drs A. Blumenthal and Z. Werb for encouragement, Dr W. J. Bodell for providing unpublished manuscript on repair of alkylation damage, and Mr G. H. Thomas and W. Charles for technical assistance. This work was supported by the US ERDA.

JAMES E. CLEAVER

Laboratory of Radiobiology,  
University of California,  
San Francisco, California 94143

Received 15 July; accepted 3 October 1977.

1. Noll, M. *Nature* **251**, 249–251 (1974).
2. Olins, A. L., Carlson, R. D., Wright, E. B. & Olins, D. E. *Nucleic Acids Res.* **3**, 3271–3291 (1976).
3. Oudet, P., Gross-Bellard, M. & Chambon, P. *Cell* **4**, 281–300 (1975).
4. Weintraub, H. *Cold Spring Harbor Symp. quant. Biol.* **38**, 247–256 (1973).
5. Searle, R. L. *Cell* **9**, 423–429 (1976).
6. Piper, P. W. *et al.* *Nucleic Acids Res.* **3**, 493–505 (1976).
7. Wilkins, R. J. & Hart, R. W. *Nature* **247**, 35–36 (1974).
8. Mortelmans, K., Friedberg, E. C., Slor, H., Thomas, G. H. & Cleaver, J. E. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2757–2761 (1976).
9. Cleaver, J. E. in *Molecular Human Cytogenetics ICN/UCLA Symposium* (in the press).
10. Cleaver, J. E. *Adv. Radiat. Biol.* **4**, 1–75 (1974).
11. Cleaver, J. E. *Meth. Cancer Res.* **9**, 123–165 (1976).
12. Bodell, W. J. *Nucleic Acids Res.* **4**, 2619–2628 (1977).
13. Sciudero, D. & Strauss, B. *Mutat. Res.* **35**, 311–324 (1976).
14. Patrick, M. H. & Gray, D. M. *Photochem. Photobiol.* **24**, 507–513 (1976).
15. Cleaver, J. E. *Nature* **218**, 652–656 (1968).
16. Regan, J. D., Setlow, R. B. & Ley, R. *Proc. natn. Acad. Sci. U.S.A.* **68**, 708–712 (1971).
17. Paterson, M. C., Lohman, P. H. M. & Sluyter, M. L. *Mutat. Res.* **19**, 245–256 (1973).
18. Cleaver, J. E. *Photochem. Photobiol.* **12**, 17–28 (1970).
19. Cleaver, J. E., Thomas, G. H., Trosko, J. E. & Lett, J. T. *Expl Cell Res.* **74**, 67–80 (1972).
20. Smith, C. A. & Hanawalt, P. C. *Biochim. biophys. Acta* **447**, 121–132 (1975).
21. Cleaver, J. E. *Proc. natn. Acad. Sci. U.S.A.* **69**, 428–435 (1969).
22. Haynes, R. H. *Radiat. Res. Suppl.* **6**, 232 (1966).
23. Gianelli, F. & Pawsey, S. A. J. *Cell Sci.* **15**, 163–176 (1974).
24. Ahmed, F. E. & Setlow, R. B. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1548–1552 (1977).
25. Cleaver, J. E. & Bootsma, D. A. *Rev. Gen.* **9**, 19–38 (1975).
26. Kraemer, K. H., Andrews, A. D., Barrett, S. F. & Robbins, J. H. *Biochim. biophys. Acta* **442**, 147–153 (1976).

## Anisotropy of the Young's modulus of bone

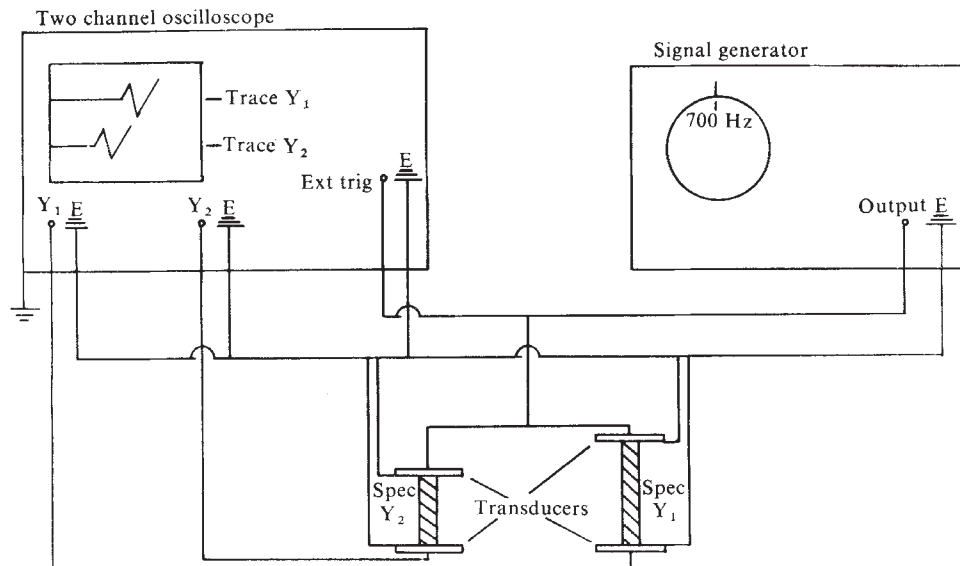
THE anisotropy in elastic deformation exhibited by compact bone sections is well known<sup>1</sup>, but has been demonstrated mainly for orientations parallel to, (designated as longitudinal) or, normal to, (designated as transverse or radial), the long axis of the bone. Various models based on the concept that bone is a hydroxyapatite-reinforced collagen composite have been proposed (for example ref. 2) to account for the elastic behaviour of compact bone, as defined by the Young's modulus ( $E$ ), but lack of data for a range of orientations has precluded one critical test of their validity. We report here preliminary measurements on the Young's modulus of mature bovine femur cortical bone for specimens orientated at various angles from 0° to 90° to the bone long axis. These results were obtained on a series of rectangular-shaped, machined, specimens (4 mm × 3 mm) with two different lengths (15 mm and 20 mm), by a novel differential ultrasonic technique, which is illustrated in Figs 1 and 2.

The time between the generation and receipt of a signal was measured (in  $\mu$ s) for specimens of 15 mm and 20 mm length, and as all other features of the circuit remained constant, the difference in time ( $dt$ ) for the two conditions was related to the difference in path length ( $d$ ) provided by the bone specimens, and the velocity ( $v$ ) through the bone calculated. Assuming the path length in bone is isotropic, the Young's modulus ( $E$ ) is then derived from:—

$$\frac{dl}{dt} = v = \left( \frac{E}{\rho} \right)^{\frac{1}{2}} \quad (1)$$

where  $\rho$  is the specific gravity of bone<sup>1</sup>. This procedure was repeated for pairs of specimens orientated at various angles to the longitudinal axis (in a tangential rotation).

The absolute values of  $E$  calculated for the longitudinal (0°) orientation (17.3–19.7 GNm<sup>-2</sup>) are within the range of values (7–27 GNm<sup>-2</sup>) obtained by various workers<sup>1</sup> from the measurement of stress-strain curves, although smaller than a single value obtained on 'dry' bovine femur sections by an ultrasonic technique (26 GNm<sup>-2</sup>) (ref. 3). For the transverse orientation (90°), the calculated values of  $E$  (8.7–10.3 GNm<sup>-2</sup>) compare with values of 7.2 GNm<sup>-2</sup> (ref. 1) (stress-strain curve) and 13.2 GNm<sup>-2</sup> (ref. 4) (shock tube loading). The present results indicate that  $E_0/E_{90} = 2.3$  (dry) or 1.7 (wet), which agrees reasonably with previous findings<sup>1</sup> of a ratio of 1.9 (wet). These points of agreement suggest that the results obtained by the present technique are valid.

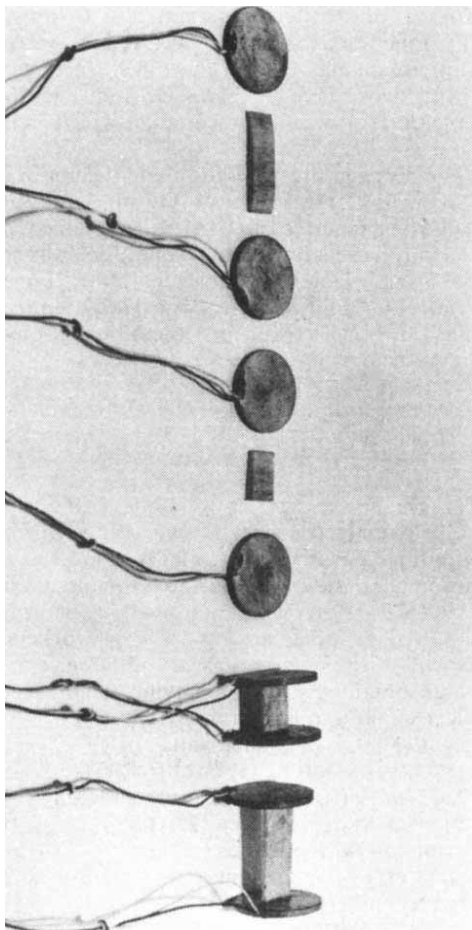


**Fig. 1** Circuit arrangement, with a signal through two specimens  $\gamma_1$  and  $\gamma_2$  producing traces  $Y_1$  and  $Y_2$ , respectively, on a two-channel oscilloscope.

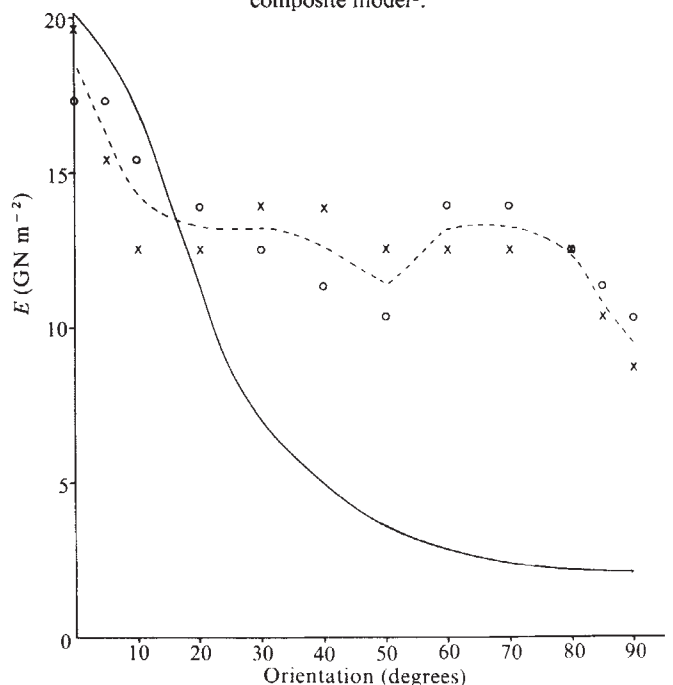
No other results have been reported for the Young's modulus of bovine femur specimens orientated at angles of 5, 10, 20, 40, 50, 70, 80 and 85° to the long axis, which are shown in Fig. 3. The Young's modulus values at 0, 30, 60 and 90° compare with values of 23.1, 16.7, 12.8 and 10.4  $\text{GNm}^{-2}$ , respectively, measured by Reilly and Burstein<sup>5</sup>.

The significant finding of the present study is the experimentally-determined variation in Young's modulus with orientation, which exhibits a decrease in  $E$  from 0–20°, an approximate plateau from 20–70° (with a 'low' point at 50°) and another decrease from 70–90°. This result is in conflict with the predictions of a fibre-reinforced composite model, such as demon-

**Fig. 2** Details of the experimental arrangement of transducers and bone specimens, before and after mounting. (Specimens shown are not of the experimental dimensions.)



**Fig. 3** Variation in Young's modulus of bovine femur specimens ( $E$ ) with the orientation of specimen axis to the long axis of the bone, for wet (○) and dry (×) conditions compared with the theoretical curve (—) predicted from a fibre-reinforced composite model<sup>2</sup>.



strated by the theoretical curve<sup>2</sup> (derived for  $E_{0^\circ}/E_{90^\circ} = 10$ ) plotted for comparison in Fig. 3, which reveals a progressive reduction in the dependence of Young's modulus on orientation for values of  $E$  from  $\approx 10^\circ$  to  $90^\circ$ , with an almost constant value of  $E$  in the 70–90° range.

We conclude, therefore that an alternative model is required to account for the dependence of Young's modulus on orientation. Further work is in progress to extend these results to a series of bones with different ultrastructures.

This research forms part of a project on the deformation and fracture of bone supported by the SRC.

W. BONFIELD  
M. D. GRYPAS

Department of Materials,  
Queen Mary College,  
London E1, UK.

Received 24 June; accepted 14 September 1977.

- 1 Evans, F. G. *Mechanical Properties of Bone* (Charles C. Thomas, Springfield, Illinois, 1972).
- 2 Currey, J. D. *J. Biomech.* **2**, 477–480 (1969).
- 3 Lang, S. B. *Science* **165**, 287–288 (1969).
- 4 Bonfield, W. & Datta, P. K. *J. Mat. Sci.* **9**, 1609–1614 (1974).
- 5 Reilly, D. & Burstein, A. H. *J. Biomech.* **8**, 393–405 (1975).

# reviews

## Nuclear phenomena and development

A. T. Sumner

*Nuclear Cytology in Relation to Development.* By F. d'Amato. Pp. viii+283. (Cambridge University: New York, London and Cambridge, 1977.) £15.

THE illustration on the jacket of this book of Balbiani rings on a polytene chromosome seems particularly appropriate for the subject of *Nuclear Cytology in Relation to Development*. In fact, Professor d'Amato takes his subject matter from a much wider field than this illustration might suggest. "Nuclear cytology" is regarded as ranging from molecular biology to cytogenetics; what one might regard as nuclear cytology in the strict sense—the detailed study of nuclear structure—does not occupy an important place in this book. Similarly, the subject of "development" is broadly interpreted. Evidently, although this is not stated in the preface, Professor d'Amato's aim has been to describe those nuclear phenomena which are associated with the many aspects of development. A refreshing and often illuminating aspect of his treatment of the subject is the balance of examples from both the plant and animal kingdoms.

The first chapter deals with life-cycles, especially the alternation of diploid and haploid phases, and certain aspects of fertilisation. Chapter 2, on the cell cycle, opens with a brief excursion into molecular biology, discussing DNA sequence complexity, as well as the different classes of RNA. It is disappointing to find that a book published in 1977 should regard histones as "general repressors of gene activity", while making no mention of their important structural role in nucleosomes. This section seems rather oddly placed at the beginning of a chapter very much concerned with factors influencing the duration of different stages of the cell cycle, although also treating various aspects of chromosome replication. From the cell cycle, d'Amato turns to a well integrated account of meiosis, with special consideration of meiosis in relation to parthenogenesis.

It is not at all clear why chapter 4,

on "Mosaics and Chimaeras" was included. Although the subject is of great relevance to several problems of development, no special nuclear phenomena are described here, apart from a short mention of chromatin elimination in cases of gonadosomic mosaicism. This chapter highlights a problem which recurs to a varying extent throughout the book. There are many excellent and well described examples in every chapter, but insufficient attempt has been made to explain their relevance to the chapter in which they occur, or to the book as a whole. Since, as already noted, the author does not clearly state his aim in writing the book, it is particularly important that a common theme should be evident throughout the book. Such a theme eventually appears as one reads the book, but there are places where it seems that the author's enthusiasm for certain topics has led to considerable digressions.

The remaining half of this book sticks more closely to what seems to be its main theme: nuclear phenomena correlated with development. Two chapters deal with changes in nuclear DNA content during development: one with increases of the whole chromosome complement (essentially polyploidy and polyteny), and the other with differential DNA replication. There is much of interest in both these chapters, and the author's practice of taking examples from a wide variety of organisms is of particular advantage here. Evidently, DNA amplification, either selective or total, may be more important in differentiation than I, for one, had supposed.

Next come two chapters which cover what one might regard as the essence of a study of the nucleus in relation to development: gene expression and its regulation. It is axiomatic, as d'Amato clearly states, that differentiation must be viewed in terms of gene expression. Although none of the nuclear changes described earlier in this book can be causally related to differentiation, it is clear that single somatic nuclei of plants and (as described in the final chapter) of animals

contain all the genetic information necessary for normal development. d'Amato's consideration of gene expression concentrates on the localisation by *in situ* hybridisation of repeated genes, and on transcription in lampbrush and polytene chromosomes. These latter aspects could perhaps have been treated more extensively with advantage, for here are two types of chromosomes which afford unrivalled material for the study of nuclear cytology in relation to development. It is, moreover, a section which calls for half-tone illustrations, of which there are none in the book. If these were rejected on grounds of expense, it seems a false economy. (The book is nevertheless well illustrated by line drawings.) In the following chapters, the roles of heterochromatin, histones, and non-histone chromosomal proteins in gene regulation are considered, with the growing evidence for the importance of the latter being clearly brought out. The final chapter, "Regeneration and Totipotency," brings together information interesting in its own right, including the exciting experiments on nuclear transplantation in amphibia. This work, showing that nuclei of differentiated somatic cells still contain all the information necessary to produce a complete adult organism, is now such a cornerstone of developmental biology that it deserves an earlier and more prominent place in this book.

In conclusion, Professor d'Amato has brought together a wealth of information on the changes which occur in cell nuclei in different developmental situations. This is reflected in a bibliography of (the publisher tells us) over 1100 references, occupying nearly 50 pages. One wishes that the author had taken more space for general comment to integrate his material. Nevertheless, this is a book in which the specialist may well see new light shining on old problems, and the non-specialist read profitably for general information. □

A. T. Sumner is a member of the scientific staff of the MRC Clinical and Population Cytogenetics Unit, Edinburgh, UK.



## Electron spectroscopy

*Electron Spectroscopy: Theory, Techniques and Applications*. Vol. 1. Edited by C. R. Brundle and A. D. Baker. Pp. xv+459. (Academic: New York and London, 1977.) £24; \$46.90.

A MAJOR EFFORT to survey the present situation in electron spectroscopy has been undertaken by some of its more prominent explorers under the editorship of Drs C. R. Brundle and A. D. Baker. The material is distributed in three volumes, the first of which is now ready. To get a balanced impression of the total work one would have to await the other two volumes, since much basic and important material is still left to be treated in these forthcoming volumes. Already, after going through the first volume, one is, however, in a good position to see that this work will become an indispensable source of knowledge for all investigators in the field.

After an introduction to electron spectroscopy by the two editors, the present many-electron theory of photoelectron emission is reviewed by R. L. Martin and D. A. Shirley. The authors have themselves contributed a good deal to some of the conceptual parts of the theory for electron emission processes. After reading this and the next chapter by W. L. Jolly—treating his equivalent cores approximation—the reader should have obtained a good understanding of some of the main topics, which underlay most applications. W. C. Price, well known as an authority in ‘classical’ ultraviolet photon spectroscopy for molecules before photoelectron spectroscopy had appeared and also as one who has given distinct impetus to present-day photoelectron spectroscopy, treats the latter as applied to small molecules. Particularly interesting is his discussion as to why previous emission and absorption photon spectroscopy failed to reproduce the molecular electron level systems. He concludes his chapter thus: “In the brief period in which it has been developed, photoelectron spectroscopy has had phenomenal success in revealing the electronic structure of matter in a particularly direct way. Chemists can see the orbital structure of even fairly large molecules and no longer have to rely on the predictions of theoreticians. Future advances in the study of solids, adsorbed species, etc., can confidently be predicted to yield information on the nature and functions of the electrons involved. It is also clear that the subject will provide a happy hunting ground

for the physicist as well as the chemist for many years to come.”

It is gratifying to see that indeed many chemists are already on the hunting ground and with most rewarding results. One of the great hunters is no doubt E. Heilbronner who, together with his colleague J. P. Maier in the next chapter, gives many illustrative examples from his rich scientific experience in the field of organic chemistry. A corresponding account from inorganic chemistry is amply given by R. L. DeKock, written, according to the editors’ preface, during his stay in Beirut during “the troubled period of civil strife in Lebanon!”

High temperature studies of ultraviolet-excited electron spectra from

halides is treated by one of the pioneers in this advanced form of the spectroscopy, J. Berkowitz, and the volume ends with a chapter on two-parameter coincidence experiments by M. E. Gellender and A. D. Baker. This technique, as pointed out by the authors, is more tedious than the usual electron spectroscopy but is instead applicable to, for example, ions of chosen excitation energies.

The applications of electron spectroscopy are expanding rapidly and already there is a problem to keep abreast with achievements in various fields. The present high quality volume is most helpful in this respect.

**Kai Siegbahn**

*Kai Siegbahn is Professor in the Fysiska Institutionen, Uppsala, Sweden.*

## World food supply problem

*World Food Resources: Actual and Potential*. By Michael Allaby. Pp.vii+418. (Applied Science: London, 1977.) £15.

THIS ‘popular’, readable, fact-packed book is presumably aimed at the ‘informed layman’. But much of it should interest the scientist, agriculturalist or economist who wishes to see his speciality in better perspective. After an introductory chapter, comes a useful chapter (2) on population and on (economic as distinct from biological) food demand, followed by two chapters on regional nutritional status, the cost of ‘modernising’ farming, the ‘green revolution’, soils, water and fertilisers (3 and 4). Two historical chapters (5 and 6) are too UK-based to be useful. There then follow helpful chapters (7–9) about ‘advanced’ farming and its dependence on non-solar energy inputs, on its genetic vulnerability, on its ecological side-effects, and about alternative farm technologies.

An interesting chapter (10) on fish is followed by one (11) on land reform, rural employment, farm credit, and so on, and the *possible* (reviewer’s italics) means of raising incomes and food production. Nevertheless, “we cannot be certain that those most in need of food will receive it” (p357). The final chapters (12 and 13) are politico-economic and examine the likelihood that effective international trade, aid and monetary devices would help to secure food for the under-fed, for example, by making phosphatic fertiliser available to poor countries.

Allaby, if not pessimistic, is doubtful about the likelihood of positive international action (chapter 13). But such doubts, one feels, are challenges to

those scientists and agriculturalists who are, for instance, seeking techniques to make available to crop plants the phosphate that is now immobilised in the soil, or to raise the lysine content of maize or sorghum, or to devise more productive farming systems (for example, the work at the International Institute of Tropical Agriculture, Ibadan, Nigeria). Allaby fails, it seems, to appreciate the scope for increasing, by relatively inexpensive means, the efficiency of human food chains. A major function of those trying to raise such efficiency should be, one feels, to make food production and consumption in poor countries as biologically independent as possible both of income gradients and of international squabbles, ideological differences, economic collapses and weather disasters.

The book as a whole lacks editorial discipline: chapter titles could be less ‘slick’ and more descriptive; more subtitles in the text would help the reader; not all the many tables are necessary although some of them, it must be said, are not available elsewhere in printed form. The references are inconsistently presented (some on tables, some at chapter ends, some in the Bibliography) and at least one is missing. Finally, many of the frequent acronyms are not defined or are enigmatic except to those who just happen to know, for example, that FAO/IWP means the Indicative World Food Plan (IWP) issued by the United Nations Food and Agricultural Organisation (FAO) in 1970.

This is a readable and useful, but sometimes diffuse, confusing and perhaps unduly pessimistic contribution on the world food supply problem. A much shorter, more disciplined version would be of greater use to the scientific community. **A. N. Duckham**

*A. N. Duckham is Emeritus Professor of Agriculture at the University of Reading, UK.*

## Evolution updated

*Evolution.* By T. Dobzhansky, F. J. Ayala, G. L. Stebbins and J. W. Valentine. Pp. 572 (Freeman: San Francisco and Reading, 1977.) £11.60.

EACH co-author has written four chapters in this book, setting forth his own ideas, and reflecting his own involvement with biology and genetics. I have a feeling that this has inevitably, although inadvertently, resulted in some overlap, not necessarily of illustrative material, but in the broad outlines of the ground covered. As might be expected, the most lucid passages are by Dobzhansky, to whose memory the book is dedicated. A notable 'one-liner' by him, "Nothing in biology makes sense except in the light of evolution", is on a flyleaf following the title page.

The chapter titles are as follows: "The Nature of Evolution", "Patterns of Speciation", "Evolution of Prokaryotes and Unicellular Eukaryotes", and "The Future of Evolution" by Stebbins; "The Genetic Structure of Populations", "The Origin of Hereditary Variation", "Phylogenies and Macromolecules", and "Philosophical Issues" by Ayala; "Natural Selection", "Populations, Races, Sub-species", "Species and Their Origins", and "Evolution of Mankind" by Dobzhansky; "Transspecific Evolution", "The Geological Record", "Cosmic Evolution and the Origin of Life", and "The Evolutionary History of Metazoa" by Valentine. These widely ranging topics certainly provide ample basis for a sweeping treatment of evolution.

Molecular matters have apparently been assigned to Ayala, who is a population geneticist. He also describes his studies of polymorphism in *Drosophila* populations of various localities. Valentine is a geologist and palaeobiologist, and he has provided an excellent discussion of the fossil record of evolutionary rates, and the record of ancient life and environment, including a welcome section on plate tectonics. Stebbins and Dobzhansky are both well known as authors of books and treatises on evolution.

In commenting on the 'evolutionary clock', Ayala discusses (p309) the comparison of amino acid differences between  $\alpha$ -haemoglobins of mammals and carp. He says that "the most recent ancestor to the four-legged mammals lived some 70 million years ago", and therefore he concludes that the comparison is useful for only 10% of the 700 million years of evolution of the  $\alpha$  chain. But this contention overlooks the fact that  $\alpha$ -haemoglobin sequences are also available for kangaroo, echidna, chicken,

viper and newt, and these fall in the intervening years. Ayala (p309) notes that differences between  $\alpha$ -haemoglobins of man and mice (17) are less than between rabbits and mice (28). (These differences are actually 16 and 27.) From this, he speculates that the rates of amino acid replacement may be increased by shorter generation time. The data he cites, however, are too scanty. If the  $\beta$  chains are compared, the man-mouse difference is 28 and the rabbit-mouse difference is 29.

The subject of evolution embraces both the past and future of human beings. It therefore presents an irresistible temptation to indulge in philosophical discourse. Three of the authors have availed themselves of this opportunity, and Ayala holds forth on "Philosophical Issues" for the final 42 pages, in which he provides us with such insights as "Common sense tells one that children resemble their parents and that good seeds produce good crops", and "No organism can be truly independent of the environment". I felt that Stebbins, and, more especially, Dobzhansky, had provided sufficient (and excellent) treatment of dissertational matters in chapters 14 and 15, so that the last chapter was not needed.

In a comparison of cytochrome sequences, the table on pages 296 and 297 shows only 20 identical positions shaded in grey, but 36 positions should have been so indicated. In addition, the

sequence of residues 27–34 in *Neurospora* cytochrome is erroneous.

Figure 2-3, page 26, depicts the anticodors for UUU and GGU as AAA and ACC. These should be GAA and GCC. Figure 3-1 shows an erroneous codon for methionine, and represents purines by "Y", which is the symbol for pyrimidines.

All in all, the book is an excellent and up-to-date treatment of evolution, the subject that has spread like a network to encompass all the biological sciences. I could not help contrasting the wealth of material in this book with the fact that creationists are still largely successful in their efforts to minimise the teaching of evolution in schools in the USA. In June 1977, a member of the California State Board of Education was able to expunge or garble certain sentences descriptive of evolution in the forthcoming new edition of *Science Framework*, a booklet of guidelines for teaching science. For, as Dobzhansky says (p439): "The mutually sustaining effects of biologists, paleontologists and anthropologists have made the theory of the evolutionary descent (or rather, ascent) of man impregnable . . . And yet some antievolutionists persist. Some of them are simply ignorant of the evidence, while others have so prejudged the question that no evidence is meaningful to them".

Thomas H. Jukes

Thomas Jukes is Professor of Medical Physics at the University of California at Berkeley.

## Botanical heritage

*Dictionary of British and Irish Botanists and Horticulturists: Including Plant Collectors and Botanical Artists.* By Ray Desmond. Pp. xxvi+747. (Taylor and Francis: London, 1977.) £40.

If there is one area in which the British Isles has been highly productive, it is in the generation of botanists, yet few of their names have survived except possibly in the specific names of certain obscure plants. Names like Ray, Gerard, Turner and Culpeper may still be familiar to many, especially those caught up in the recent general enthusiasm for health foods, herbs, and so on. But many more names have now passed into oblivion as botanical exploration and taxonomy have progressed. This *Dictionary* has grown out of a desire to catalogue our British and Irish botanical heritage and to collate what information can be recovered concerning the achievements of our deceased, plant-hunting forefathers.

Each entry is set out in the form of details of date and place of birth and death, education, qualification and

selected publications. Information concerning biographies, obituaries, herbaria and species graced with the botanist's name are also given in abbreviated form. Only dead botanists are included, but all spheres of the subject are covered, from nurserymen to palaeobotanists and physiologists.

The obvious question one must ask when faced with this hefty and expensive tome, concerns the use to which the book may be put. It is not a collection of historical anecdotes, and will not, therefore, serve as a book into which one may dip for entertainment. Its value is likely to be appreciated only by those undertaking serious historical studies into the development of botany in Britain. It is a book which will thus find its way into specialist libraries, but at this price it is hardly likely to attract the attention of individual botanists. It is essentially a work of reference which will lead the historically minded to obscure sources which would not otherwise be easily traced.

Peter D. Moore

Peter Moore is Senior Lecturer in Plant Sciences at King's College, University of London, UK.



## Neutron scattering

*Neutron Scattering in Chemistry.* By G. E. Bacon. Pp. 186. (Butterworth: London and Boston, Massachusetts, 1977.) £12.50.

IN writing this book, Professor Bacon has attempted a very difficult task. It is difficult firstly because any technique is likely to be able to contribute to many different branches of chemistry and secondly because neutron scattering is several techniques rolled into one. On the one hand, it is a structural tool and so immediately embraces the whole range of structural work on condensed matter including crystallography. In addition, however, the magnetic dipole moment of the neutron makes possible the study of magnetic structures and hence the determination of unpaired spin densities. On the other hand, it is also a spectroscopic technique because the low energy of neutrons of wavelength comparable with molecular dimensions makes possible an energy analysis of scattered neutrons so as to reveal details of atomic and molecular translational, rotational and vibrational motions. It is this unique ability to probe explicitly the time-dependent structure that makes neutron scattering so powerful. The range of problems which can usefully be studied, however, is always limited by the available neutron fluxes.

For many years, it has been possible to study relatively complex structures but when the fourth dimension of energy analysis is added the flux determines the practicable energy resolution and therefore the complexity of the problem which is amenable to study. A major advance in this respect was provided by the advent of the latest generation of high flux reactors within the past few years, and in particular the unique instruments at the Institut Laue Langevin, Grenoble. It is unfortunate that the timing of this book is such that only a very little of this exciting new work could be included.

Professor Bacon (a Professor of Physics) has attempted to give a brief (~ 180 pages) and essentially qualitative description of some of the results of applying neutron scattering techniques to various areas of chemical interest. It is a book in which to find a summary of some of the things which have been or can be done using neutrons, rather than a detailed description of the underlying theory or of the chemistry. For the reader who wishes to pursue a topic to a greater depth, there is a general bibliography of neutron scattering texts and conference proceedings and a list of references with each chapter.

The first two chapters contain good, succinct descriptions, respectively, of the

Principles of Neutron Scattering and Experimental Methods. Three chapters follow on crystallographic studies (Structural Studies, Direct Methods, and Correlation of X-Ray and Neutron Data: X-N Syntheses) which contain some well-chosen examples of recent work. A brief chapter on Studies of Biological Materials contains a timely discussion of small angle scattering methods, but it is surprising to find no mention of specifically chemical applications of these techniques, such as the important studies of polymeric materials. Two further chapters complete the discussion of crystal structure investigations; the first (Measurements of Covalency) is concerned with magnetic scattering and the second with both Bragg reflection and diffuse scattering studies of Defects and Non-Stoichiometry in a variety of crystalline phases.

In all of the above, as well as in the structural aspects considered in the last chapter on Liquids, Glasses and Gases the treatment is clear and authoritative. The two chapters not so far mentioned are entitled Molecular Spectroscopy and

## Raman spectroscopy

*Raman Spectroscopy.* By D. A. Long. Pp. xiii+276. (McGraw-Hill: New York and London, 1977.) £14.40.

FOLLOWING the development of the laser in the early 1960s, there has been a boom in Raman spectroscopy. It is now a routine laboratory technique. This rapid progress has led to the publication of several multi-author volumes and to the *Journal of Raman Spectroscopy*. This book, however, is the first by a single author that aims to provide an "up-to-date survey of the whole subject, setting Raman spectroscopy in perspective, unifying the basic theory, illustrating the applications and potential of the technique, and guiding the reader towards the specialist literature".

There are eight chapters and three appendices. Also, there is a central reference section of sixteen blue pages containing thirteen tables of formulae for intensities and polarisation properties of Raman scattering by gases. Chapter 1 is a general introduction, a nice feature being the reproduction of the original 1928 papers by C. V. Raman and K. S. Krishnan and by G. Landsberg and L. Mandelstam in *Nature* and *Naturwissenschaften*, respectively. Chapter 2 describes the properties of electromagnetic radiation and defines the four Stokes parameters. Chapters 3 and 4 present classical and partial quantum-mechanical treatments of Rayleigh and Raman scattering. They proceed at an easy pace and include a very detailed account of the

Polymers. Here, the treatment is less satisfactory partly because, as mentioned earlier, it is in these areas where marked advances have very recently been made, since the availability of high resolution spectrometers. Nevertheless, this does not account for all the omissions: for example, there is no discussion of quasielastic scattering from molecular rotations, nor of solid-state diffusion; indeed, the treatment of quasielastic scattering does not begin to reflect the importance of this area of work (the omission of any reference to Springer's elegant 1972 monograph on the subject demonstrates the point). There should have been some mention of phase transitions and a more extensive discussion of phonon dispersion in molecular crystals. In spite of the omissions, however, this is a timely and welcome book which goes some way towards filling a serious need

A. J. Leadbetter

A. J. Leadbetter is Professor of Physical Chemistry at the University of Exeter, UK.

polarisability tensor. Chapter 5 is on time-dependent perturbation theory, the Placzek polarisability theory, resonance Raman scattering, and Raman optical activity. Experimental procedures are described in chapter 6 and numerous examples are discussed in chapter 7. The final chapter is devoted to non-linear Raman effects. Appendix I is a guide to the literature; II lists for the common point groups the symmetry classes of translations, rotations, polarisabilities and first hyperpolarisabilities; and III describes a Raman study of a crystal.

The style is lucid and the illustrations excellent (although the notation is not always attractive), so the book will be helpful to many students. It expounds at length some of the properties of cartesian tensors, and includes some very interesting and recent applications of Raman spectroscopy to problems in organic and inorganic chemistry. Some of the more physical topics, such as line shapes and Brillouin scattering from liquids, are not considered. In a few places, as in the description of magnetic and electric Raman optical activity on p131, there are errors; and the references are not always adequate (for example, the four references at the end of chapter 3 and those in Appendix I give no help to someone trying to understand non-linear polarisation, introduced on p41).

The book has been well produced and is good value.

A. D. Buckingham

A. D. Buckingham is Professor of Chemistry at the University of Cambridge, UK.



# obituary

## Gersh Budker

PROFESSOR Gersh Itskovich Budker, Director of the Institute of Nuclear Physics of the Siberian Branch of the Soviet Academy of Sciences, died on 5 July 1977. With his death we have lost one of the most imaginative and colourful accelerator scientists.

Professor Budker was born on 1 May 1918 in the Ukraine. He graduated from Moscow University in 1941 and, after having served in the Soviet Army until 1945, began work at the Institute of Atomic Energy of the USSR Academy of Sciences. He was initially involved in nuclear reactor development but his interests gradually moved to accelerators for elementary particle research. In 1956 he also became professor at the Moscow Engineering Physics Institute.

In the same year the Scientific Research and Educational Centre, Akademgorodok, was created just outside Novosibirsk. This was a great challenge to creative young scientists, and it was natural that Budker, in 1957, was appointed director of its Institute of Nuclear Physics. This Institute has been the source of an almost continuous stream of ideas and novel accelerator systems and very soon became known throughout the world.

In his new position Budker was able to realise some of the plans he had already developed while at the Kurchatov Institute in Moscow. His main theme was to provide colliding-beam facilities so as to bring his institute to the forefront of elementary particle physics without the heavy expenses required for more conventional accelerators. He was not only afraid of great expense (he even sold accelerators in the 'open' market to boost his budgets), but also afraid of the inflexibility inherent in very large installations. The programme he established was ambitious and daring and made heavy demands on ingenuity and drive.

He had his first storage ring, VEP-1, working in 1963. An electron-positron device, VEPP-2 of  $2 \times 700$  MeV, was in operation from 1967, another one, VEPP-3 of  $2 \times 2.2$  GeV, in 1972, and a fourth one, VEPP-4 of  $2 \times 7$  GeV, is at present under construction. They have made it possible for his institute to

Sorry, for copyright reasons some images on this page may not be available online

Camera Press

perform front-line experiments on meson resonances. Some of the rings have also been used as a source of synchrotron radiation for research in solid state physics, etc.

Budker had incorporated a two-way proton colliding beam device in the original plans for his Institute. He modified this in 1966 when he invented the so-called electron cooling of proton or antiproton beams, which made it feasible to propose instead a colliding beam project with antiprotons against protons. The electron cooling technique was essential for the accumulation of sufficiently intense beams of antiprotons.

But first the cooling principle had to be verified experimentally. With the limited means available at the institute this took longer than Budker had hoped and delayed the proton-antiproton plans. However, in 1974 his institute could report most beautiful experimental evidence for electron cooling of a proton beam. This invention and its experimental verification must be considered the peak of Budker's career. The technique and its possible uses are being discussed in accelerator laboratories all over the world, and it is being considered in important plans both at the European Organisation for Nuclear Research, CERN, and at the Fermi National Accelerator Laboratory in the United States.

Although Budker and his institute became best known because of the work on accelerator devices, the institute also has a branch that carries out research in plasma physics, which gives a very fruitful cross-fertilization between plasma physics and accelerator physics.

Professor Budker distinguished himself not only as a scientist, but also through his exceptional ability to stimulate and inspire his collaborators. His Institute is now 20 years old and has many remarkable achievements behind it. Nevertheless, it still retains an unusual pioneering spirit. Budker also felt very deeply about the whole philosophy that lay behind Akademgorodok and was a driving force in ensuring that this unique place became a centre of scientific achievements with a world-wide reputation.

Budker was a member of the USSR Academy of Sciences. In 1967 he received the Lenin Prize for his contributions to the development of colliding beam devices.

K. Johnsen

## Milton N. Bramlette

MILTON NUNN BRAMLETTE, the distinguished American geologist, died on 31 March, 1977. He was born in Bonham, Texas, on 8 February, 1896, attended preparatory school in St. Louis, and entered the University of Wisconsin, Madison, in 1914.

His university studies were interrupted by World War 1. He enlisted in the aviation service, and qualified as a pilot, but too late in the war for combat service. He was discharged early in 1919 as a 2nd Lieutenant, and returned to Madison, where he graduated A.B. in 1921. During two summers he had served as assistant on the Wisconsin Geological Survey, running magnetic traverses across the iron ranges of northern Wisconsin.

With this field experience and high academic record, he was appointed Assistant Geologist in the U.S. Geological Survey in 1921. While headquartered in Washington he commuted to Johns Hopkins for advanced work in stratigraphy. After field mapping in Montana, Kansas and South Dakota, he took leave of the Survey

for graduate work at Yale. This was interrupted by field work in Mexico and Venezuela for the Gulf Oil Company. He returned to Yale for further graduate studies in 1928–1930, when he resumed his appointment with the Geological Survey. His doctorate was received *in absentia* in 1936.

Bramlette soon entered into a close and remarkably fruitful association with that outstanding specialist in Mollusca, Wendell P. Woodring, for a series of brilliant stratigraphic studies in California: the Palos Verdes Hills, Kettleman Hills and Santa Maria Valley. It was then that Bramlette began his petrological studies of the sedimentary rocks, dealing especially with intrastratal alterations, phosphates, zeolites and siliceous cements, culminating in his classic paper on the cherts of the Monterey Formation. He also began his studies of Cenozoic micropalaontology and with his keen discernment was soon able to identify many of the diagnostic foraminifera in the field with a hand lens.

Bramlette was a principal investigator, along with W. H. Bradley and K. E. Lohman, of the transatlantic series of Piggott cores during the 1930's. With our current plethora of much longer deep sea cores, this work has been nearly forgotten; nevertheless it was the first study of deep-sea stratigraphy and was highly significant in permitting the first correlation of European and American glaciations. At this time he laid the groundwork, through his studies of the Arkansas bauxite deposits, for the great increase in productivity of these mines during the submarine-induced crisis of the second world war.

In 1940 Bramlette joined the faculty of the University of California at Los Angeles, where he remained until his transfer to the Scripps Institution of Oceanography in 1951, except for two years during World War II, when he worked on the bauxite deposits of Arkansas and Jamaica.

At the Scripps Institution Bramlette began his highly rewarding studies of the fossil coccoliths. Because of their minute size and irregular forms, these organisms sink very slowly in the sea so that currents distribute them almost world-wide. Accordingly they constitute unusually valuable time markers for the widespread correlations of marine sediments. Bramlette's contribution to the development of such correlations was pre-eminent, and he continued as a world leader long after his formal retirement in 1964.

Bramlette's versatility and profound scholarship were outstanding. He was elected to the National Academy of Sciences in 1954, was awarded its Thompson Medal in 1964, the Distinguished Service Medal of the Depart-

ment of the Interior in 1963, and the Doctor of Laws degree from the University of California in 1965.

He died of emphysema on March 31, 1977. He was a modest gentleman.

James Gilluly

## A. L. Walpole

DR ARTHUR WALPOLE died at Wilmslow, Cheshire on 2 July 1977, aged 64. Trained initially as a chemist (Imperial College London, B.Sc., Ph.D.), he spent a short period in the Department of Pharmacology, Edinburgh, and then joined the biological section of the new pioneer group in the Dyestuffs Division of ICI Ltd, in Manchester, which ultimately became the Pharmaceuticals Division.

His early interest, on the implication of novel synthetic oestrogens in the treatment of tumours, coupled with the appointment of Professor (later Sir) Alexander Haddow as a consultant, began Walpole's concern with cancer research, for which he is best known. His work fell under two headings. First came his collaboration with Michael Williams, who as Medical Officer for the Dyestuffs Division was deeply involved in its associated industrial hazards. Together over the years, with *ad hoc* synthesis where needed, they worked out structure/carcinogenic relationships, particularly amongst the arylamines, which still remain of world-wide significance.

Alongside this, and arising in part from the textile interests of the Division, Walpole was amongst the earliest (1947 onwards) to study the effect of introducing alkylating groups such as methylolamido, epoxide and aziridyl into potential tumour-inhibitory structures. His recognition of the peculiar cytotoxic properties induced by these moieties again had implications for industrial safety.

Walpole's other over-riding interest was in the hormonal control of reproduction and related phenomena. Thus researches around basic derivatives of the oestrogenic triphenylethylenes have recently led to a compound especially useful in the treatment of certain forms of mammary carcinoma. Likewise, he pioneered parallel studies in animal husbandry, first using certain disthioureas, and later the new synthetic prostaglandins. He was engaged on this latter work at the time of his death, two years after formal retirement.

As a person, he naturally attracted a wide circle of friends and collaborators, and his other qualities extended deeply

into the arts. Above all, he was much concerned for humanity, whether in the mass or the individual. He leaves a widow, Dora.

F. L. Rose

## G. G. Villela

PROFESSOR Gilberto G. Villela, the founder of modern biochemistry in Brazil, died on 17 July 1977, aged 73.

Dr Villela was born in Minas Gerais, Brazil, on 12 July 1904. After graduating in medicine in 1926 he entered the Oswaldo Cruz Institute with which he was associated, first as a student and then as a member of the staff, until the time of his death.

Dr Villela was an authority in the field of biochemistry and he made many significant contributions specially in vitaminology and enzymology. Another field of active interest was the biochromes, where he made many contributions to the study of animal pigments. His publications, more than 300 papers, appeared in medical and biochemical journals in Brazil, Britain, the United States and elsewhere. He was also the author of several scientific books.

Dr Villela visited, lectured and carried out scientific work at many important research centres all over the world, including the University of California Medical School (1945), University College London (1949), the Pasteur Institute (1950), and New York University (1961), where he was a Visiting Professor. As a UNESCO representative he set up a new centre of biological research at the University of Rangoon, Burma (1967). More recently he went to Japan as a lecturer, at the University of Nagoya (1973) and as Chairman of the Xth International Congress of Nutrition in Kyoto (1975).

Dr Villela was elected to several international scientific societies and received several medals and prizes. He was the founder and first President of the Brazilian Society of Biochemistry, and attended most of the Society's meetings.

Dr Villela's life was fully devoted to his scientific interests which he pursued with enthusiasm and vigour. To his co-workers and students he was an inspiring and ever-helpful teacher. He enjoyed discussions; his mind was very critical but combined with a good sense of humour. He was a man of high moral standing and humanistic education. He will always be warmly remembered by all who knew him.

He is survived by his widow Regina, one son, Gilberto, one married daughter, Sonia and a grandson, Pedro.

Emilio Mitidieri